

TOPIC 1: MEASUREMENT

H2 Physics Syllabus 9749

SECTION I: Measurement

Measurement	Learning Outcomes Students should be able to:
Physical quantities and SI units	(a) recall the following base quantities and their SI units: mass (kg), length (m), time (s), current (A), temperature (K), amount of substance (mol). (b) express derived units as products or quotients of the base units and use the named units listed in 'Summary of Key Quantities, Symbols and Units' as appropriate. (c) use SI base units to check the homogeneity of physical equations. (d) show an understanding of and use the conventions for labelling graph axes and table columns as set out in the ASE publication <i>Signs, Symbols and Systematics: The ASE Companion to 16–19 Science</i>.8 (e) use the following prefixes and their symbols to indicate decimal sub-multiples or multiples of both base and derived units: pico (p), nano (n), micro (μ), milli (m), centi (c), deci (d), kilo (k), mega (M), giga (G), tera (T). (f) make reasonable estimates of physical quantities included within the syllabus.
Scalars and vectors	(g) distinguish between scalar and vector quantities, and give examples of each. (h) add and subtract coplanar vectors. (i) represent a vector as two perpendicular components.
Errors and uncertainties	(j) show an understanding of the distinction between systematic errors (including zero error) and random errors. (k) show an understanding of the distinction between precision and accuracy. (l) assess the uncertainty in a derived quantity by addition of actual, fractional, percentage uncertainties or by numerical substitution (a rigorous statistical treatment is not required).

SPECIAL NOTE

This set of notes contains QR codes to direct you to relevant Youtube videos or websites. To make full use of this feature, please download a free **QR code reader app** onto your smartphone.

If you are using a desktop, you may visit <https://goo.gl/lob7Bp> to access the Youtube playlist for this chapter.

You are strongly encouraged to view the videos especially on the worked solutions, and complete the respective examples.

Contents

TOPIC 1: MEASUREMENT	2
1.1 Introduction	4
1.2 Physical Quantities and Units	4
1.2.1 Base Quantities and Units	5
1.2.2 Derived Quantities and Units	5
1.2.3 Homogeneity of equations	7
1.2.4 Prefixes	9
1.2.5 Estimation	9
1.3 Errors and Uncertainties	11
1.3.1 Significant Figures and Decimal Places	13
1.3.2 Combining Uncertainties	14
1.3.3 Precision and Accuracy	18
1.3.4 Systematic and Random Errors	20
1.4 Scalars and Vectors	26
1.4.1 Adding and Subtracting Coplanar Vectors	26
1.4.2 Resolution of Vectors	31
Further Reading and Reference	35
Appendix I International System of Units (SI)	35
Appendix II Summary of Key Quantity, Symbols and Units	36
Appendix III Homogeneous Equation	38
Appendix IV SI Prefixes	39
Appendix V Rules of Significant Figures in Mathematical Operations	40

1.1 Introduction

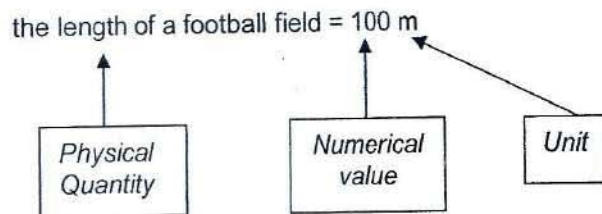
Measuring our observations objectively is an important task that differentiates science from the arts. The scientific method involves making observations that are repeatable by another person. For this reason, the study of physical quantities and their units is the starting point of our learning of science.

1.2 Physical Quantities and Units

A physical quantity is a physical property that can be measured and/or calculated and expressed in numbers. For example, "length" is a physical quantity that can be expressed by stating a number of a measurement unit such as metres or inches, while "beauty" is a property that is difficult to describe with a number.

Hence **the value of a physical quantity** is expressed as **a numerical value and a unit of measurement**.

For example,



This has a few implications.

1. Using the unit "metre" helps us to define a particular property of the football field – its length. Using other units could allow us to define other properties (for example, square metres describe the area) of a football field.
2. We can only compare properties of the same unit. Using the "metre" allows us to compare the length of one object with another. We cannot compare a "metre" with a "kilogram".
3. The numerical value of the physical quantity is dependent on the unit it is expressed in. The length of the football field will have a value of 330 when expressed in the unit "foot".

Further Reference:

Explore Science Stories on the topic Measurement

(<http://sciencelearn.org.nz/Science-Stories/Measurement>)



1.2.1 Base Quantities and Units

Through the centuries, scientists in different parts of the world have used different units at different times to measure the same physical quantities. However standardised definitions and units are needed in the modern world.

The current International System of Units^a, universally abbreviated SI (from the French *Le Système International d'Unités*), is the modern metric system of measurement.

Seven well-defined and independent base quantities are chosen. Their corresponding SI base units are indicated:

Base Quantity		Base Unit	
Name	Symbol	Name	Symbol
Mass	m	kilogram	kg
Length	l	metre	m
Time	t	second	s
Electric current	I	ampere	A
Thermodynamic temperature	T	kelvin	K
Amount of substance	n	mole	mol
Luminous intensity	I_v	candela	cd

Note:

1. The units, when written in full, are in small letters. Ampere and Kelvin are names of scientists.
2. The kilogram is the only SI unit with a prefix as part of its name and symbol.
3. Luminous intensity is not in the LO(a) of the Syllabus.
4. Definitions of the units are not required in the Syllabus but are provided in Appendix I at the end of the lecture notes for your reference.

IMPORTANT TO KNOW

SI base units were chosen based on their mutual independence. They are the simplest units that cannot be expressed in terms of other SI base units.

1.2.2 Derived Quantities and Units

There are a lot of other physical properties that we come across in our daily lives. These quantities and units used scientifically are included in the SI^b as derived quantities and units. Note that these derived units, such as newton and joule, are also SI units.

IMPORTANT TO KNOW

Derived quantities are related to the base quantities through mathematical and scientific equations. Derived units (the units of derived quantities) are defined in terms of the base units using the same equations.

One simple example is the derived quantity of area. It is defined using the mathematical equation $\text{area} = \text{length} \times \text{breadth}$, where both length and breadth are defined in terms of the base quantity, length. Hence the unit for area can be written in terms of base units as m^2 .

^a The SI was established in 1960 by the 11th General Conference on Weights and Measures (CGPM, Conférence Générale des Poids et Mesures). (<http://www.bipm.org/en/si/>)

^b There are some commonly used units that are not in the SI. Some examples are the minute, hour, day, degree, litre, tonne, electron-volt, unified atomic mass unit, nautical mile, knot, hectare, bar and angstrom.

Some derived units have special names and symbols. An example is the newton (symbol N), a unit for force. We can define this derived quantity - force, using formulae such as $F = ma$, where F is force, m is mass and a is acceleration. You may note that acceleration itself is also a derived quantity. Using such an equation, we can then find the representation of the newton (N) in terms of base units (kg m s^{-2}).

Study the examples in the following table and verify the **derive units** for the respective quantities.

Derived quantities, symbols	Formula	Derived units	Usual units
Volume, V	-	m^3	-
Density, ρ	$\rho = m/V$	kg m^{-3}	-
Velocity, v	$v = s/t$	m s^{-1}	-
Acceleration, a	$a = (v - u)/t$	m s^{-2}	-
Force, F	$F = ma$	kg m s^{-2}	N
Momentum, p	$p = mv$	kg m s^{-1}	N s
Pressure, p	$p = F/A$	$\text{kg m}^{-1} \text{s}^{-2}$	Pa
Energy, E	$KE = \frac{1}{2} mv^2$	$\text{kg m}^2 \text{s}^{-2}$	J
Power, P	$P = E/t$	$\text{kg m}^2 \text{s}^{-3}$	W
Electric charge, Q	$Q = It$	A s	C
Voltage, V	$V = W/Q$	$\text{kg m}^2 \text{s}^{-3} \text{A}^{-1}$	V
Resistance, R	$R = V/I$	$\text{kg m}^2 \text{s}^{-3} \text{A}^{-2}$	Ω
Frequency, f	$f = 1/T$	s^{-1}	Hz

For more information on the symbols and units that will be used in question papers, refer to Appendix II.

Example 1.1

According to Newton's Law of Gravitation, the force F between two point masses M and m separated by the distance r is given by the formula $F = \frac{GMm}{r^2}$, where G is the universal gravitational constant.

Obtain the SI base units for G . [$\text{m}^3 \text{s}^{-2} \text{kg}^{-1}$]

$$\begin{aligned} G &= \frac{Fr^2}{Mm} \\ &= \frac{\text{N} \cdot \text{m}^2}{\text{kg} \cdot \text{kg}} = \frac{\text{kg} \cdot \text{m} \cdot \text{s}^{-2} \cdot \text{m}^2}{\text{kg} \cdot \text{kg}} \\ &= \text{m}^3 \text{s}^{-2} \text{kg}^{-1} \end{aligned}$$

Watch detailed worked solution:

**Unitless (or dimensionless) quantities**

- all numbers, e.g. 2, $\frac{1}{2}$, π , e
- trigonometrical functions, e.g. sine, cosine, tangent
- all logarithmic functions, e.g. $\log_{10}$, $\ln$
- powers, e.g. 10^y , the ratio $\frac{x}{y}$ must be unitless. If x has a unit, then the unit of y must have the same unit as that of x .
- Unitless physical constants: e.g. refractive index of glass, relative density of a liquid.

1.2.3 Homogeneity of equations

For an equation to be physically plausible, the equation must be *homogenous*. This is to say the units on both sides of the equations must be the same. This is because only quantities with the same units may be equated, added or subtracted.

Example 1.2

Bernoulli's equation, which applies to fluid flow states that

$$P + h\rho g + \frac{1}{2}\rho v^2 = k \quad \text{where } P \text{ is pressure, } h \text{ is height, } \rho \text{ is density, } g \text{ is acceleration due to gravity, } v \text{ is velocity and } k \text{ is a constant.}$$

Show the LHS of the equation is homogeneous and state the SI unit for k . [$\text{kg m}^{-1} \text{s}^{-2}$]

$$\text{Unit of } P = \text{Nm}^{-2} = (\text{kg m s}^{-2})(\text{m}^{-2}) = \text{kg m}^{-1} \text{s}^{-2}$$

$$\text{Unit of } h\rho g = \text{m}(\text{kg m}^{-3})(\text{m s}^{-2}) = \text{kg m}^{-1} \text{s}^{-2}$$

$$\text{Unit of } \frac{1}{2}\rho v^2 = (\text{kg m}^{-3})(\text{m s}^{-1})^2 = \text{kg m}^{-1} \text{s}^{-2}$$

Since all terms have the same unit, the LHS of the equation is dimensionally consistent (shown).

Watch detailed worked solution:



$$\text{SI unit of } k = \text{kg m}^{-1} \text{s}^{-2}$$

Checking of homogeneity of an equation is not only a powerful way of establishing if an equation is reasonable. It also provides hints for guessing the actual equation.

Example 1.3

Consider the period T of a simple pendulum. The possible factors which may affect it are its length l , its mass m and the acceleration due to gravity g . Use unit analysis to arrive at a plausible formula for T .

$$T: l, m, g$$

$$T = k l^a m^b g^c$$

$$\text{units of LHS} = s^1$$

$$\text{units of RHS} = m^a kg^b (ms^{-2})^c$$

$$= m^{a+c} kg^b s^{-2c}$$

$$b = 0$$

$$-2c = 1 \Rightarrow c = -\frac{1}{2}$$

$$a + c = 0 \Rightarrow a = \frac{1}{2}$$

$$T = k l^{\frac{1}{2}} g^{-\frac{1}{2}} \Rightarrow T = k \sqrt{\frac{l}{g}}$$

Watch detailed worked solution:



While an equation which is not homogeneous must be wrong, an equation which is homogeneous may not be correct (see Appendix III). Ultimately, the validity of an equation can only be verified through experiments.

Further reference:

MIT OpenCourseWare Video – the Art of Approximation



Prefixes

The rationale for prefixes is simple. While it is alright to write your height as 1.65 m, it will be quite cumbersome to write the width of your hair as 0.00005 m, or the size of an atom as 0.0000000001 m. In mathematics, you have learnt the use of standard form. The size of an atom can be rewritten more neatly as 1×10^{-10} m but scientists often use prefixes instead of standard form for values between 10^{-12} and 10^{12} . Prefixes can be used with both base units and derived units.

Factor	Prefix	Symbol	Name	Decimal equivalent	Order of magnitude
10^{-12}	pico	p	Trillionth	0.000,000,000,001	-12
10^{-9}	nano	n	Billionth	0.000,000,001	-9
10^{-6}	micro	μ	Millionth	0.000,001	-6
10^{-3}	milli	m	Thousandth	0.001	-3
10^{-2}	centi	c	Hundredth	0.01	-2
10^{-1}	deci	d	Tenth	0.1	-1
10^0	-	-	One	1	0
10^3	kilo	k	Thousand	1,000	3
10^6	mega	M	Million	1,000,000	6
10^9	giga	G	Billion	1,000,000,000	9
10^{12}	tera	T	Trillion	1,000,000,000,000	12

The accepted convention is to use a prefix such that the quantity can be written as a whole number and of least significant figures (s.f.). For example the wavelength of red light would be written as 650 nm, rather than 0.65 μ m. Values larger than the ranges listed in the prefixes above should be written in standard form (The above list of prefixes are stipulated in LO(e). Do not use the other prefixes listed in Appendix IV).

You may notice that the symbol for the prefix milli (m) is the same as that for the base unit metre (m), leading to a confusion when we see, for example, “ms⁻¹”. It could mean “metre per second”, or “per millisecond”. To distinguish the units especially in print, the A-level standard is to use “ms⁻¹” for “per millisecond” and leave a space “m s⁻¹” for the derived unit “metre per second”.

Further reference:

Powers of Ten video



The Scale of the Universe (<http://htwins.net/scale2/>)



1.2.5 Estimation

When was the last time you estimated something? How much does that shirt at G2000 cost? How long does it take to travel by public transport from Hwa Chong to Suntec City? If your 1-litre water bottle were to leak one drop of water at a time, how much time will it take for the water bottle to empty? There is a myriad of questions which we can answer by means of measurement and calculation. We could measure the average speed of a bus and the distance along the bus route and calculate the travelling time. We could also estimate.

Physicists frequently use “back-of-the-envelope” calculations or “Fermi” problems, named after Physicist Enrico Fermi who worked on the Manhattan Project during World War II. Fermi was known for making approximate calculations with little or no actual data. One well-documented example was his estimate of the strength of the atomic bomb based on the distance travelled by bits of paper dropped from his hand at the test blast^c.

Estimation in Physics uses simple numbers (1, 2 or 5) with the correct order of magnitude (10^3 or 10^{-4}). It is important to know whether a lecture theatre can sit a few students, tens, hundreds or thousands of students, but not so important to know that it has a capacity of 327. When the exact answer or values are not known, assumptions and estimation are used to find a rough answer.

Estimation is also very useful as a check to the answer that we would get from complex calculations.

Examples of “Fermi” problems:

- Estimate the total number of hairs on your head.
- How many bricks are needed to build a home?
- Estimate the efficiency of an electric kettle.
- How much paper is used by Hwa Chong Institution for lecture notes in 1 year?

Some important estimates that you should memorize:

- Typical wavelength of visible light: 10^{-7} m (range: 400 to 700 nm)
- Range of wavelengths of other electromagnetic regions
- Size of an atom: 10^{-10} m or 1 angstrom (Å)
- Size of a nucleus: 10^{-15} m

Estimation problems are also popular interview questions for tech companies.^d

Strategy for estimation:

1. Identify the unknown: Define specifically what you need to estimate
2. Identify the known: Find your experience that may help you relate to the unknown
3. Find a relationship between the known and the unknown: make a connection between what you know and what needs to be estimated.

Example 1.4 [N08/1/2]

Which estimate is realistic?

- A The kinetic energy of a bus travelling on an expressway is 30 000 J.
 B The power of a domestic light is 300 W
 C The temperature of a hot oven is 300 K.
 D The volume of air in a tyre is 0.03 m^3 .

Watch detailed worked solution:



A: $KE = \frac{1}{2}mv^2$
 Assuming bus travels at 60 km/h
 $= \frac{1}{2}(10000)(\frac{60000}{3600})^2 = 1.4 \times 10^6 \text{ J}$

B: $5 \text{ W} - 100 \text{ W}$ C: $300 \text{ K} = 26.85^\circ\text{C}$

D: Vol. of air = $[\pi R^2 - \pi r^2] \times h$
 $= [\pi (0.30^2) - \pi (0.20^2)] \times 0.15$
 $= 0.024 \text{ m}^3$

^c http://en.wikipedia.org/wiki/Fermi_problem

^d <http://www.glassdoor.com/index.htm>



1.3 Errors and Uncertainties

No measurement is perfectly accurate or exact^o. **Uncertainty of measurement is the doubt which exists about the result of any measurement.** You might think that well-made rulers, clocks and thermometers should be trustworthy, and give the right answers. However, for every measurement - even the most careful - there is always a margin of doubt.

Many instrumental and human limitations cause measurements to deviate from the actual or true value of the physical quantities being measured.



The length of this card could be measured as 8.5 cm but this is surely not the exact true value of its length.

The *experimental error* in measuring a physical quantity can be understood as the difference between the *measured value* and the *true value* of a physical quantity:

$$\text{experimental error} = \text{measured value} - \text{true value}$$

The tricky part of finding the experimental error is that we usually do not know what the true value is! (which is why we were measuring it in the first place.) While we can try to improve on the measurement process, we can never eliminate experimental errors entirely. Indeed a measurement is of little use if we do not know how large its experimental error is.

The true value of a quantity to be measured may never be known. The next best thing a physicist can do in order to know how close his measured value is to the true value, is to *estimate* the experimental error in a measurement by determining the uncertainty of the measurements. The *uncertainty* in measuring a physical quantity is reflected by the *variation or differences* in the *measured values* during the process of measurement.

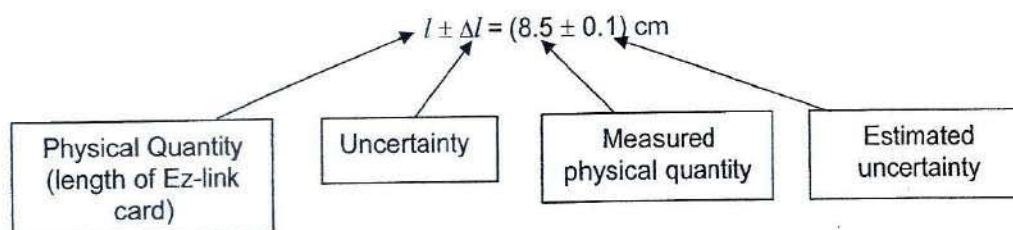
For example, the length of the Ez-link card, l , can be written as $l \pm \Delta l = (8.5 \pm 0.1)$ cm because the smallest division on the scale of the ruler is 0.1 cm.

Any position on the ruler can be read off to a position that is the closest half a division, (within 0.05 cm). Given that the length of the Ez-link card is in fact the difference in two measurements on the ruler (the left edge and right edge positions of the card),

- the largest possible length is obtained by taking the largest possible reading of the right edge position minus the smallest possible reading of the left edge position.
For example, $8.55 \text{ cm} - (-0.05 \text{ cm}) = 8.60 \text{ cm}$.

^o Reference website: <http://www.lhup.edu/~dsimanek/errors.htm>

- the smallest possible length is obtained by taking the smallest possible reading of the right edge position minus the largest possible reading of the left edge position.
For example, $8.45 \text{ cm} - 0.05 \text{ cm} = 8.40 \text{ cm}$
- Therefore, the measurement of length l is stated as:



The **actual (or absolute) uncertainty** in this measurement of length is said to be $\Delta l = 0.1 \text{ cm}$.



There are many ways to estimate the uncertainty. The example concerning the Ez-link card is one of the simplest method, which is based on the maximum possible range of readings. However, in 'A'-level written papers, the uncertainty of a particular measurement is usually given in the question, and you need not estimate it. Always use the given uncertainty values and do not replace them.

Fractional uncertainty

The **fractional uncertainty** is the ratio of the actual uncertainty to the measured value.

The fractional uncertainty in a quantity x is $\frac{\Delta x}{x}$.

Example: The fractional uncertainty of l , $\frac{\Delta l}{l} = \frac{0.1}{8.5} = 0.012$

Percentage uncertainty

The **percentage uncertainty** is obtained by converting the fractional uncertainty into percentage form by multiplying by 100%,

The percentage uncertainty in a quantity x is $\frac{\Delta x}{x} \times 100\%$

Example: The percentage uncertainty of l , $\frac{\Delta l}{l} \times 100\% = \frac{0.1}{8.5} \times 100\% = 1.2\%$.



1.3.1 Significant Figures and Decimal Places

The general rules are:

1. Express uncertainties to 1 s.f.

If the uncertainty in a measurement is estimated to be 0.025 for example, it should be rounded to 0.03.

Remember that an uncertainty is only an *estimate* of the difference between a measured value of a quantity and its true value. It makes little sense to place emphasis on the smaller deviations (0.005) as compared to 0.03.



Note however that fractional and percentage uncertainties should not be rounded off to 1 s.f. They are usually left to 2 or 3 s.f.

2. Express the quantity to the same place value as its uncertainty.

Using the same example, 0.03 has its most significant digit in the second decimal place. Hence the quantity should be written to exactly two decimal places:

$$12.10 \pm 0.03$$

and not 12.1 ± 0.03 or 12.102 ± 0.03 .

Refer to Appendix V for more useful sig. fig. rules in mathematical operations.

Example 1.5 [J84/II/1 modified]

A student makes measurements from which he calculates the speed of sound to be 327.66 m s^{-1} . He estimates that the percentage uncertainty is 3%. Round off the speed to an appropriate number of significant figures. [330 m s^{-1}]

$$\frac{\Delta v}{v} \times 100\% = 3\% = 0.03$$

$$\Delta v = 0.03 \times 327.66 = 10 \text{ m s}^{-1} \text{ (1sf)}$$

$$v \pm \Delta v = (330 \pm 10) \text{ m s}^{-1} \text{ (v rounded to same value as } \Delta v)$$

$$\therefore v = 330 \text{ m s}^{-1}$$

Watch detailed worked solution:



1.3.2 Combining Uncertainties

There are established statistical rules for calculation of consequential uncertainty from individual pieces of information. The A-level course only requires a simplified version of the statistical treatment.



(i) Adding or Subtracting Measured Quantities

When two measured quantities are added together or one subtracted from another, the actual uncertainty in the result is equal to the sum of the actual uncertainty of the two quantities.

We always add up the uncertainties even if the equation involves subtraction because we do not know the sign of the actual uncertainty and thus need to estimate the *largest possible* uncertainty in our quantities.

Rule 1: For $c = a + b$, the actual uncertainty in c , $\Delta c = \Delta a + \Delta b$
For $d = a - b$, the actual uncertainty in d , $\Delta d = \Delta a + \Delta b$



(ii) Multiplying or Dividing Measured Quantities

When two measured quantities are multiplied together or one divided by the other, the fractional uncertainty in the result is equal to the sum of the fractional uncertainties of the two quantities.

Rule 2: For $p = ab$, the fractional uncertainty in p , $\frac{\Delta p}{p} = \frac{\Delta a}{a} + \frac{\Delta b}{b}$
For $q = \frac{a}{b}$, the fractional uncertainty in q , $\frac{\Delta q}{q} = \frac{\Delta a}{a} + \frac{\Delta b}{b}$

Here, percentage uncertainty can be used interchangeably with fractional uncertainty as they differ only by a factor of 100 which is multiplied equally to every term in the equation. The percentage uncertainty in p is equal to the sum of the percentage uncertainties in a and b ,

$$\frac{\Delta p}{p} \times 100\% = \frac{\Delta a}{a} \times 100\% + \frac{\Delta b}{b} \times 100\%.$$

Let us use a simple equation to illustrate the analysis of most mathematical or scientific equations we will encounter. Consider the linear equation $y = mx + c$, where m , x and c are all measurements with their associated uncertainties. Our experimental result, y , will also have an uncertainty that depends on that of m , x and c .

The addition is simple: $\Delta y = \Delta(mx) + \Delta c$ but for the product mx , $\frac{\Delta(mx)}{(mx)} = \frac{\Delta m}{m} + \frac{\Delta x}{x}$. Rearranging,

$$\Delta(mx) = \left(\frac{\Delta m}{m} + \frac{\Delta x}{x} \right) (mx) \text{ and hence } \Delta y = \left(\frac{\Delta m}{m} + \frac{\Delta x}{x} \right) (mx) + \Delta c.$$

KIASU
ExamPaper

Example 1.6

The measurements of the dimensions of a particular piece of rectangular cardboard are (18.5 ± 0.5) mm and (12.5 ± 0.5) mm. Determine the area of the cardboard with its associated uncertainty.

$[(2.3 \pm 0.2) \times 10^2 \text{ mm}^2]$

$$\text{Area, } A = L \times W = 18.5 \times 12.5 = 231.25 \text{ mm}^2$$

Fractional uncertainty of the area,

$$\frac{\Delta A}{A} = \frac{\Delta L}{L} + \frac{\Delta W}{W}$$

$$\frac{\Delta A}{231.25} = \frac{0.5}{18.5} + \frac{0.5}{12.5}$$

$$A \pm \Delta A = (230 \pm 20) \text{ mm}^2$$

$$= (2.3 \pm 0.2) \times 10^2 \text{ mm}^2$$

("A" same place value as ΔA)

Watch detailed worked solution:



$$\Delta A = 15.5 \approx 20$$

(iii) Scaling

Actual uncertainty in a measured quantity is *scaled together* with the measured quantity.

Rule 3: If $r = ka$, then $\Delta r = k(\Delta a)$ where k is a constant

E.g. if $r = \frac{1}{4}b$ then $\Delta r = \frac{1}{4}(\Delta b)$

Think of the above as a special case of Rule 1. For instance, $r = 3a = a + a + a$

For the case of dividing measurement by a numerical constant k , think of it as multiplying the inverse, $1/k$.

Note that the fractional uncertainty $\frac{\Delta r}{r}$ is equal to $\frac{\Delta a}{a}$, as the constant k is cancelled out in the ratios.

Example 1.7

The radius of a circle is $r = (3.0 \pm 0.2)$ cm. Find the circumference with its uncertainty. $[(19 \pm 1) \text{ cm}]$

Let the circumference of circle be C .

$$C = 2\pi r = 2\pi(3.0) = 18.850 \text{ cm}$$

$$\Delta C = 2\pi \Delta r = 2\pi(0.2) = 1.26 \text{ cm} = 1 \text{ cm (to 1sf)}$$

$$C \pm \Delta C = (19 \pm 1) \text{ cm} \quad (C \text{ rounded to same place value as } \Delta C)$$

Watch detailed worked solution:

**(iv) Powers**

This applies to all exponents (n), both larger and smaller than 1, whether integer or fraction. However if n is negative, the error is still considered as positive.

Rule 4: If $s = a^n$, then $\frac{\Delta s}{s} = |n| \frac{\Delta a}{a}$

E.g. If $s = b^{-1/2}$, then $\frac{\Delta s}{s} = \frac{1}{2} \frac{\Delta b}{b}$

Example 1.8

Given a sphere of radius $r = (18.5 \pm 0.5)$ mm, find the volume of the sphere with its associated uncertainty.
 $[(2.7 \pm 0.2) \times 10^4 \text{ mm}^3]$

$$V = \frac{4}{3}\pi r^3 = 26522 \text{ mm}^3$$

Fractional uncertainty of V ,

$$\frac{\Delta V}{V} = 3 \left(\frac{\Delta r}{r} \right)$$

$$\frac{\Delta V}{26522} = 3 \left(\frac{0.5}{18.5} \right)$$

$$\Delta V = 2000 \text{ mm}^3 \text{ (to 1 sf)}$$

$$\begin{aligned} \text{Thus } V \pm \Delta V &= (27000 \pm 2000) \text{ mm}^3 \\ &= (2.7 \pm 0.2) \times 10^4 \text{ mm}^3 \end{aligned}$$

Watch detailed worked solution:

**Example 1.9** [J78/II/2; N81/II/5 modified]

In an experiment, the external diameter d_1 and internal diameter d_2 of a hollow tube are found to be (64 ± 2) mm and (47 ± 1) mm respectively. Calculate the thickness of the tube and the associated uncertainty. What is the corresponding percentage uncertainty? $[(9 \pm 2) \text{ mm}; 22 \text{ \%}]$

$$\text{Thickness } t = \frac{1}{2}(d_1 - d_2) = \frac{1}{2}(64 - 47) = 8.5 \text{ mm}$$

$$\begin{aligned} \text{Actual uncertainty in } t, \Delta t &= \frac{1}{2}(\Delta d_1 + \Delta d_2) \\ &= \frac{1}{2}(2 + 1) = 2 \text{ mm (1 s.f.)} \end{aligned}$$

$$\text{Hence, } t \pm \Delta t = (9 \pm 2) \text{ mm (same place value)}$$

$$\begin{aligned} \text{Corresponding percentage uncertainty in } t &= \frac{2}{9} \times 100 \% \\ &= 22 \% \end{aligned}$$

Watch detailed worked solution:



Example 1.10

The period of oscillation of a simple pendulum is given by $T = 2\pi\sqrt{\frac{l}{g}}$. A student conducts an experiment to find the acceleration of free fall, g . He measures the length of the pendulum, $l = 0.23 \pm 0.01$ m and the time for 20 oscillations, $t = 19.24 \pm 0.01$ s. Find g and its associated uncertainty. $[(9.8 \pm 0.4) \text{ m s}^{-2}]$

$$g = \frac{4\pi^2 l}{T^2} = \frac{4\pi^2 l}{\left(\frac{t}{20}\right)^2} = \frac{1600\pi^2 l}{t^2} = \frac{1600\pi^2 (0.23)}{(19.24)^2} = 9.81 \text{ m s}^{-2}$$

Watch detailed worked solution:



$$\frac{\Delta g}{g} = \frac{\Delta l}{l} + 2 \frac{\Delta t}{t} = \frac{0.01}{0.23} + 2 \frac{0.01}{19.24} = 0.0445$$

$$\Delta g = 0.0445 \times 9.81 = 0.4 \text{ m s}^{-2} \text{ (to 1sf)}$$

$$g \pm \Delta g = 9.8 \pm 0.4 \text{ m s}^{-2}$$

(v) Special functions or complicated functions, Z

If measurements are put together with trigonometric functions, exponential functions or other complicated formulae, use simple numerical substitutions to evaluate the uncertainty directly. The following rule may be applied.

Rule 5:

$$\begin{aligned} \text{Actual uncertainty of } Z &= \frac{1}{2} (\text{maximum possible } Z - \text{minimum possible } Z) \\ \Delta Z &= \frac{1}{2} (Z_{\text{max}} - Z_{\text{min}}) \end{aligned}$$

Example 1.11

Consider $S = x \cos \theta$ for $x = (2.0 \pm 0.2)$ cm, $\theta = (53 \pm 2)^\circ$. Find S with its uncertainty. $[(1.2 \pm 0.2) \text{ cm}]$

$$S = x \cos \theta$$

$$S = 2.0 \cos 53^\circ = 1.20$$

To get S_{max} ,

$$x + \Delta x = 2.2$$

$$(\theta - \Delta\theta) = 51^\circ$$

$$S_{\text{max}} = 2.2 \cos 51^\circ$$

$$= 1.3845 \text{ cm}$$

To get S_{min} ,

$$(x - \Delta x) = 1.8$$

$$(\theta + \Delta\theta) = 55^\circ$$

$$S_{\text{min}} = 1.8 \cos 55^\circ$$

$$= 1.0324 \text{ cm}$$

$$\Delta S = (S_{\text{max}} - S_{\text{min}}) / 2$$

$$= 0.2 \text{ (1sf)}$$

$$\text{Hence } S = (1.2 \pm 0.2) \text{ cm}$$

Watch detailed worked solution:



Summary of Rules for Consequential Uncertainties of Quantities derived from Measured Values $A \pm \Delta A$ and $B \pm \Delta B$

1. If $R = mA + nB$, then $\Delta R = |m|\Delta A + |n|\Delta B$

2. If $R = kA^m \times B^n$, where k is a numerical constant, then $\frac{\Delta R}{R} = |m|\frac{\Delta A}{A} + |n|\frac{\Delta B}{B}$

3. Other functions: e.g. $R = \sin A$, $R = \ln B$

Use the general approach: $\Delta R = \frac{1}{2} (R_{\max} - R_{\min})$

1.3.3 Precision and Accuracy



These are two terms which we need to understand in our critical approach to measurements.

Precision refers to the closeness of a set of measurements.

Measurements are said to be precise when *repeated* measurements remain very close to one another. If a certain measurement which is done, repeated several times, produces widely varying readings, the precision is poor.

For example, three measurements of my height: 1.75 m, 1.76 m, 1.75 m are precise but 1.72 m, 1.78 m, 1.76 m are not precise.

Accuracy refers to the agreement between the measured value and the true or accepted value of a quantity.

For example, if we are told that an object has a true mass 500 g, a measured value of 400 g is inaccurate. Similarly, the measurement of the acceleration of free fall, $g = 9.40 \text{ m s}^{-2}$ is also inaccurate, compared to $g = 9.70 \text{ m s}^{-2}$.

Example 1.12 Check your understanding

Watch detailed worked solution:



Complete the diagram below.

Target	Precise	Accurate
	X	X
	X	✓
	✓	X
	✓	✓

Further reference:

Rules for Calculating Uncertainties ~xmphysics0 (Summary and examples)



Proof of Product Rule (beyond the syllabus) ~xmphysics0



What's the difference between accuracy and precision? - Matt Anticole TED-Ed





1.3.4 Systematic and Random Errors

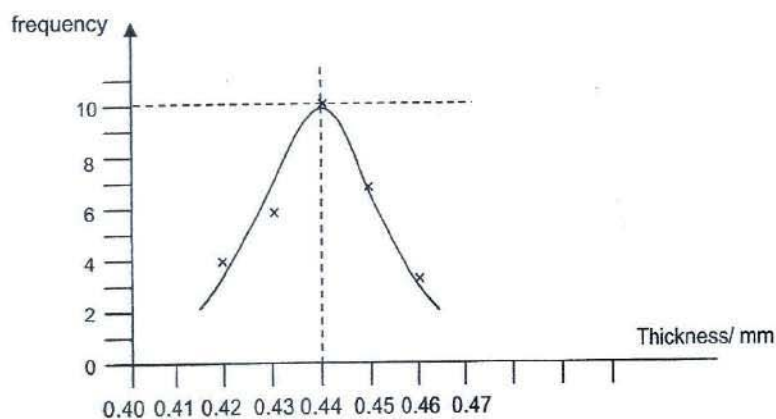
We have asserted that the true and exact value of a physical quantity can never be known. This is because our measurement of a physical quantity will always be limited by experimental error. The causes of these experimental errors can be split into two broad categories – systematic errors and random errors.

1.3.4a Random Error

Suppose we take many measurements to determine the diameter of a thin piece of wire using a micrometer screw gauge and tabulate our measurements in a table.

Readings / mm	Frequency (no. of times of occurrence)
0.42	4
0.43	6
0.44	10
0.45	7
0.46	3
Total no. of readings	30

The distribution of the measurements can be represented in the plot as shown.



Random errors occur as a *scattering* (or *spreading*) of readings about the **average** or **mean** value of the measurements. The magnitude of random errors is not fixed and each reading has equal chance of being smaller or larger than the mean value. Random errors occur despite repeating the experiments under the same conditions. This means that each reading you take will likely be different from the previous one and you cannot predict what the next reading will be.

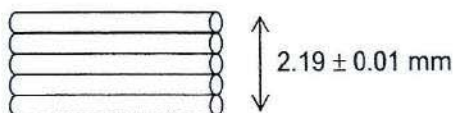
Random errors are caused by environmental conditions, irregularity of quantity being measured, and limitations of the measurement equipment or the observer.

Reducing Random Error*Repeating measurements*

While readings will appear to differ if only two or three readings were taken, the mean value of a large number of readings will tend to stabilise. This is because statistically over a large number of readings, the positive random errors tend to cancel out the negative ones hence achieving a stable average value.

Combining measurements

Instead of measuring the diameter of only one thin wire, an improvement in the experimental procedure can be done by lining up a few thin wires and stick them together side-by-side using scotch tape. We will then measure the combined diameter.



Now if we use the same micrometer screw gauge, the actual uncertainty of our measurement is still the same as $\pm 0.01 \text{ mm}$. What about the diameter of one thin wire? Since we measured the combined diameter of five thin wires, we can write $5d = (2.19 \pm 0.01) \text{ mm}$. Hence $d = \frac{1}{5}(2.19 \pm 0.01) = (0.438 \pm 0.002) \text{ mm}$. The actual uncertainty in our result has decreased tremendously and the diameter can now be expressed to one more significant figure than before.

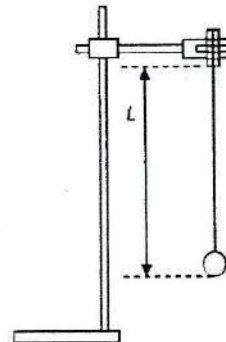
The improved procedure seems to yield a better result but there are some limitations to this method as well. Firstly the apparatus: micrometers are accurate to 0.01 mm but their range is limited to a few centimetres. Secondly the quantity: real-life wires always suffer discrepancies during manufacturing hence it is of little meaning if mathematically we can calculate the result to a very small error but in practice differences between wires or parts of the same wire is much larger.

Graphical method is also very important and widely used because experiments usually demand more than just measuring one quantity but to investigate and understand relationships between two or more physical quantities. The simplest way of doing this will be to plot a graph and draw a best-fit line. We will deal more with graphing during your practical sessions.

Example 1.13: Oscillations of a pendulum

The period of one oscillation of the pendulum T is related to the length of the pendulum L according to the equation $T = 2\pi\sqrt{\frac{L}{g}}$, where $g = 9.81 \text{ m s}^{-2}$. A diagram of the experiment is shown on the right. T is measured for one oscillation with a stop watch and L is measured with a metre ruler held in hand.

L/cm	T/s (1 st reading)	T/s (2 nd reading)
22.0	1.16	0.98
23.5	1.07	1.23
25.1	1.15	1.19
25.6	1.30	1.10
27.3	1.42	1.20
28.1	1.38	1.25



Using the given information,

- identify a random error in this experiment and explain the source of the error
- suggest an appropriate method to reduce the random error you have identified

Watch detailed worked solution:

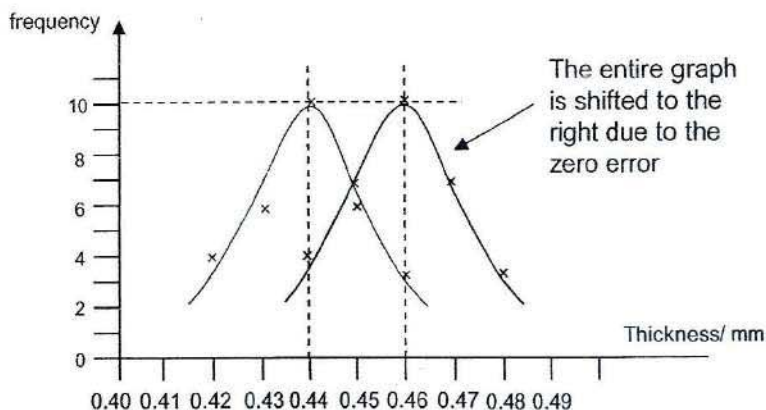


- The variation in T for a particular L is very large. The uncertainty in T is caused by the variation in human judgement of when to start and when to stop timing the oscillation. Note that the human reaction error is around 0.2s
- Time should be taken for a large number of oscillations. If N oscillations were taken, the period is calculated as $T = t/N$. We usually take N such that each t reading is more than 20s, so that the percentage error contributed by human reaction time error is less than 10% (0.2/20).

1.3.4b Systematic Error

Suppose now we discover that the micrometer screw gauge used has a zero error of +0.02 mm. Every single measurement is off by 0.02 mm.

The distribution of the readings would look like this:



This shift of values from the real values is the result of a *systematic error*. Systematic errors are reproducible errors which cause a set of readings to deviate in a fixed direction from the true value. Systematic errors are caused by instrumental errors, environmental conditions, and poor experimental techniques. The source of systematic errors can be determined and eliminated by corrective action.

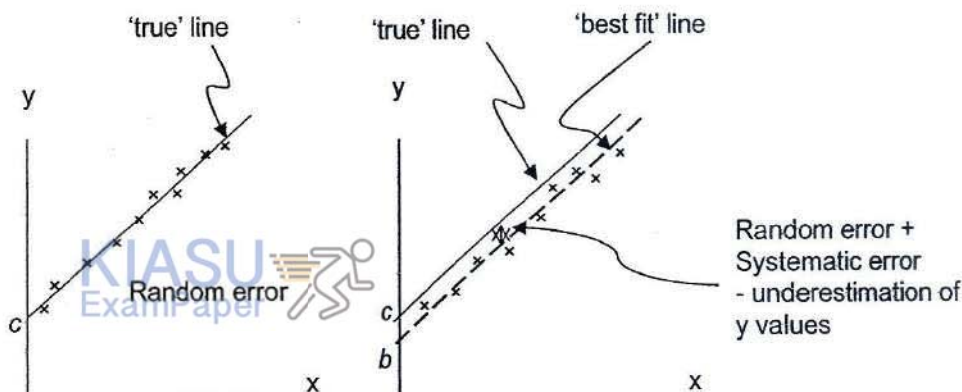
Question: Distinguish random errors from systematic errors.

Answer: Random errors are deviations of the measured value from the mean value, with varying signs and magnitudes. **Systematic errors** are deviations of the measured value from the true value, with same sign and similar magnitude.

Eliminating Systematic Error

In some cases, systematic error can be accounted for and corrected after it has been detected. The easiest way to detect systematic error is to note the vertical-intercept of the graph that is plotted. For example the graph of $y = mx + c$ when plotted gives a y-intercept of b instead of c as shown in the diagram below. The systematic error of $(c - b)$ results in an underestimation of the y -values.

The effects of random errors and of systematic errors appear in graphs as illustrated below.



Proper usage of measuring instruments is essential in eliminating systematic error. Measuring instruments need to be checked before usage. Always ensure balances, callipers, scale pointers, digital devices all read zero before being used. Calibration of instruments should be cross-checked if possible by comparing a measurement using two measuring devices.

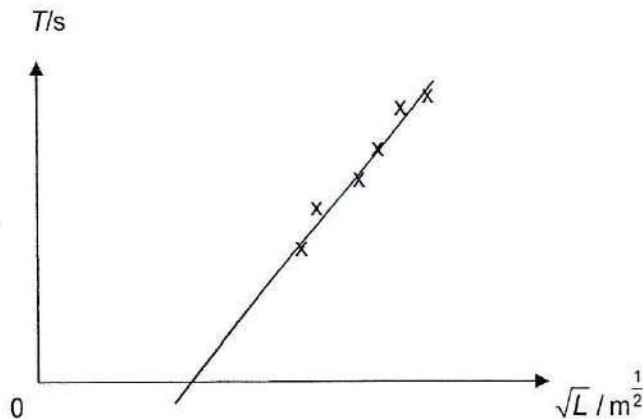
In many experiments, the most glaring error is the experimenter himself. Carelessness while taking measurements can result in severe systematic errors and random errors. However, blunders such as misreading a ruler or calculation mistakes should NOT be quoted as examples of errors in practical reports or exams!

Examples of Systematic Errors

Descriptions of errors	Error Sources	Corrections
- zero errors on the scales of instruments - poor calibration of instruments	Due to apparatus	- correct all measured readings by negating the error accordingly. - calibrate the instrument properly before experiment
consistent parallax error which affect all the readings in the same way, for instance, taking readings off a scale from a fixed angle.	Due to poor experimental technique	adopt the correct way to take reading: ensure that the <i>line of sight</i> is <i>perpendicular</i> to the measuring scale.
<i>background radiation</i> causes the count rate of your radioactive sample to be consistently higher than the true reading.	Due to external factors	Take the external factor(s) into account and adjust all readings appropriately. For instance, Measure the average background count rate and subtract it from the measured count rate.

Example 1.14: Oscillation of a pendulum

Refer to the Example 1.13 for the same setup. The graph of T against $\sqrt{L}$ is shown below:



Watch detailed worked solution:



Using the given information,

- c) identify a systematic error in this experiment and explain the source of the error.
 d) suggest an appropriate method to eliminate the systematic error you have identified.

c) The graph of T against $\sqrt{L}$ should pass through the origin. For every value of T , the length L appears to be longer than it should be.

1. The length of the pendulum could be measured wrongly.
2. The ruler was held in the hand and may not be parallel to the string during measurement

d) 1. In the experimental setup diagram, L should be measured only up to the centre of mass of the bob, not to the edge of the bob.

2. The ruler should be suspended with a retort stand. A spirit level should be used to ensure that the ruler is vertical (since the pendulum will also rest

vertically when it is stationary).

1.3.4c Summary Table

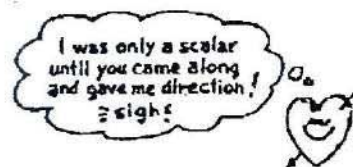
Type of error	Characteristics	If this error exists, will <u>precision</u> be affected?	If this error exists, will <u>accuracy</u> be affected?
random	varying in both magnitude and direction about a mean value, can be reduced by taking average of repeated readings, but not eliminated.	Yes	No
systematic	varying in fixed direction about a true value, CANNOT be reduced by taking average of repeated readings, but may be eliminated by good experimental techniques.	No	Yes



1.4 Scalars and Vectors

Physical quantities that have *no direction* associated with them are known as *scalars*. Scalars are specified completely by values and units. Examples of scalars include distance, speed, mass, time, temperature, energy, gravitational and electric potentials.

Physical quantities that have both a magnitude and a *direction* are known as *vectors*. Examples of vectors include displacement, velocity, acceleration, force, gravitational and electric field strengths.

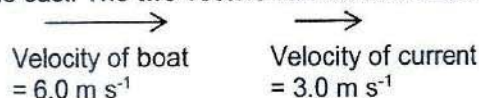


In books, a vector is often denoted in one of the following ways: $\vec{F}$, $\overline{F}$ or $\underline{f}$. The magnitude is indicated as $|\vec{F}|$ or simply F .

The direction of vectors should always be clearly presented. In written form, we can express, for example, velocity as 5 m s^{-1} towards the east, or acceleration of free fall as 9.81 m s^{-2} downwards. This is somewhat troublesome to write when we come to diagonal motions.

Diagrams are always helpful in Physics when dealing with vector quantities. Each vector is represented by an arrow. The arrow is always drawn pointing in the direction of the vector quantity and the length of the arrow is proportional to the magnitude of the vector quantity. If more than one vector is drawn on a diagram, the lengths of the different arrows should be representative of the relative magnitudes of the vector quantities.

For example, a boat's engine can propel the boat to move at 6.0 m s^{-1} to the east and the current in the river is flowing at 3.0 m s^{-1} also to the east. The two vectors can be drawn as shown.



Note the relative lengths of the arrows: the velocity of the boat is twice that of the current, hence its arrow's length is also twice as long.

Multiplication of a vector by a scalar simply scales the length of the arrow. Multiplication of a vector by -1 (or a negative sign, "-") reverses the direction in which the vector points.

1.4.1 Adding and Subtracting Coplanar Vectors

It is obvious that the boat moves down the river faster with the presence of a current than without. In addition the boat will in fact move slower if it is moving upstream or against the flow of current. Hence there must be an additive and subtractive effect when there are two or more vector quantities present in one situation.



The sum of two or more vectors is called the *resultant* vector. The resultant velocity vector of the boat is 9.0 m s^{-1} when flowing with the current and 3.0 m s^{-1} when flowing against the current.

As demonstrated in the diagram above, resultant vectors are always found by connecting one arrow-head to the tail of the next arrow. The resultant vector is then drawn from the tail of the very first arrow to the arrow-head of the last arrow. This summation process can be repeated step-by-step for any number of vectors.



1.4.1a Vectors in 1D

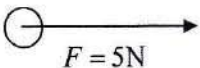
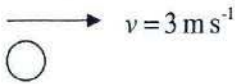
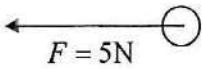
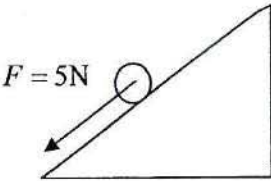
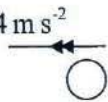
One dimensional vectors lie along a single line (collinear) and can be either pointing along or against that direction, like the current in a river. In mathematical form, we usually fix one of the directions as positive, and the opposite direction as negative.

In the direction of travel of the boat, its own velocity $\vec{v}_{\text{boat}} = +6.0 \text{ m s}^{-1}$. The velocity of river current $\vec{v}_{\text{along}} = +3.0 \text{ m s}^{-1}$ when it is flowing in the same direction as the boat's motion. If the river current is against the motion of the boat, it will be written as a negative quantity: $\vec{v}_{\text{against}} = -3.0 \text{ m s}^{-1}$. These two vectors $\vec{v}_{\text{along}}$ and $\vec{v}_{\text{against}}$ are equal in magnitude but opposite in direction $\Rightarrow \vec{v}_{\text{against}} = -\vec{v}_{\text{along}}$. The resultant velocity at which the boat travels is the *vector sum* of its own velocity and the velocity of the river current.

$$\vec{v}_{\text{resultant}} = \vec{v}_{\text{boat}} + \vec{v}_{\text{along}} \quad \text{or} \quad \vec{v}_{\text{resultant}} = \vec{v}_{\text{boat}} + \vec{v}_{\text{against}}$$

Vector values in physics

Vector values are most commonly written as numbers with a sign attached to it to indicate the direction. The **direction of the positive-value is predefined with an arrow**. For example, the vectors as shown below may be written as:

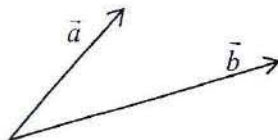
 $\rightarrow + : F = 5\text{N}$	 $\rightarrow + : v = 3\text{ms}^{-1}$
 $\rightarrow + : F = -5\text{N}$	 $\rightarrow + : v = -2\text{ms}^{-1}$
 $\nwarrow + : F = 5\text{N}$ or $\nearrow + : F = -5\text{N}$	 $\rightarrow + : a = -4\text{ms}^{-2}$ OR $\leftarrow + : a = 4\text{ms}^{-2}$



1.4.1b Vectors in 2D

Two dimensional vectors lie on the same plane (coplanar) but point in different directions that are not along a single straight line.

In the diagram below, $\vec{a}$ and $\vec{b}$ are two coplanar vectors lying on the plane of the paper but not along a single straight line.



The tail of each arrow indicates the position where a vector quantity is acting upon. In the diagram above, if both $\vec{a}$ and $\vec{b}$ are forces, they would be acting on the same point, or the same object.

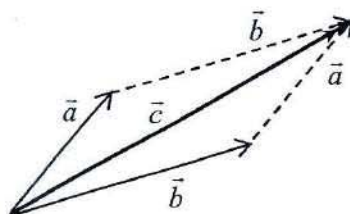
Imagine if you were pulled by two friends, one at each arm, in different directions. In which direction would you move? Can you possibly move in both directions you are pulled? There only exists one resultant force that will determine the direction you will move.

Similarly, the resultant vector $\vec{c}$ is obtained by adding vectors $\vec{a}$ and $\vec{b}$, or $\vec{c} = \vec{a} + \vec{b}$.

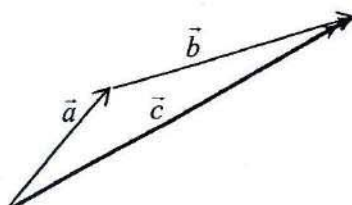
Addition of vectors: $\vec{c} = \vec{a} + \vec{b}$

Methods of Determining the Resultant Vector

- a) **Parallelogram Method:** By putting the vectors to be added 'tail to tail'. Complete the parallelogram. The resultant vector is the diagonal from the tail of the two vectors to the other vertex of the parallelogram.



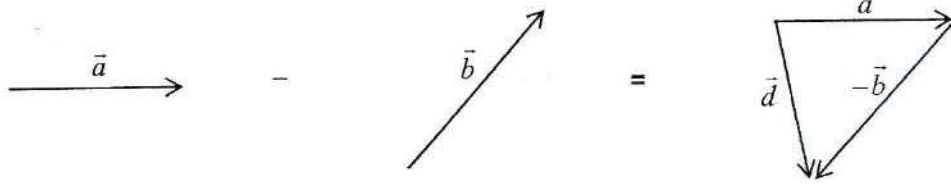
- b) **Triangle Method:** By joining the end of the next arrow onto the tip of the previous one to form a chain of arrows, the resultant vector is a straight arrow that goes from the tail of the chain directly to the head. The order of adding the vectors does not affect the resultant.



- c) **Component Method:** The vectors are resolved into two perpendicular directions and then added. Please see section 1.4.2. **This method is extremely important because it is the most convenient method to add three or more vectors.**



Subtraction of vectors: $\vec{d} = \vec{a} - \vec{b}$



In Physics, vector subtraction can come about when we want to determine the change in a certain physical vector quantity.

In general,

A change in a physical quantity = final value – initial value

Mathematically it is written as

$$\Delta Q = Q_f - Q_i$$

Physical quantities that are often seen in A-level questions include:

Scalars: change in temperature, change in energy

Vectors: change in velocity, change in momentum

Further reference:

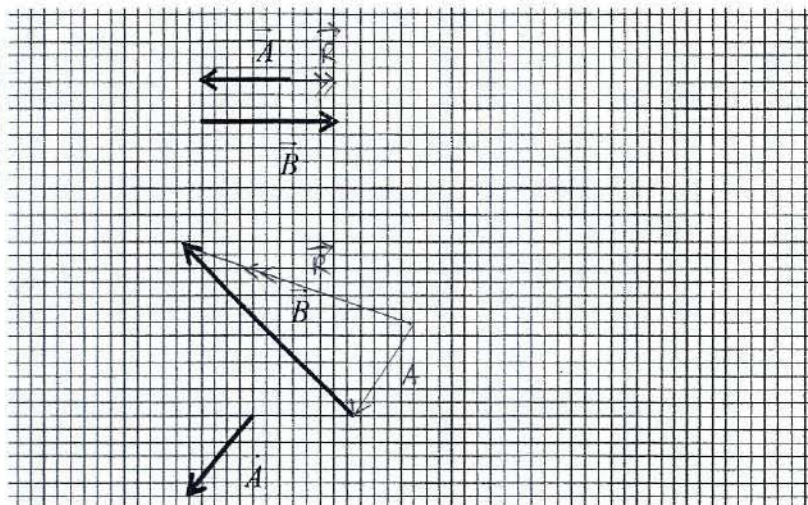
Vector Addition and Subtraction ~xmphysics0 (Summary and examples)



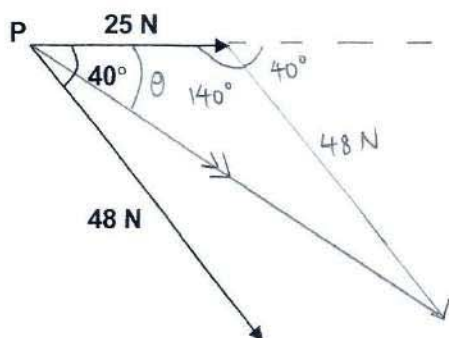
Example 1.15

(a) In the figure below, for each pair of vectors $\vec{A}$ and $\vec{B}$, draw the resultant vector $\vec{R}$ where $\vec{R} = \vec{A} + \vec{B}$.

Watch detailed worked solution:



(b) Two forces act at a point P as shown below. Determine (magnitude and direction of) the resultant of these two forces. [69 N, 26.6° clockwise from the 25 N force]



Find magnitude using cos rule,
 $R^2 = 25^2 + 48^2 - 2 \times 25 \times 48 \cos 140^\circ$
 $R = 69 \text{ N}$

Find angle θ using sin rule,

$$\frac{\sin \theta}{48} = \frac{\sin 140^\circ}{69}$$

$$\theta = 26.6^\circ$$

The resultant force is 69 N,
 26.6° clockwise from the 25 N force.

- (c) An object is moving at 5.0 m s^{-1} due east. Its direction changes to due south with a speed of 7.5 m s^{-1} . Determine (i) the change in speed. (ii) the change in velocity.
[2.5 m s^{-1} ; 9.0 m s^{-1} 33.7° west of south]

(i). Change in speed = $7.5 - 5.0 = 2.5 \text{ m/s}$

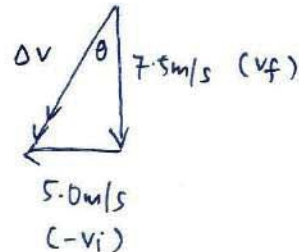
(ii). Change in velocity

$$\Delta v = v_f - v_i$$

$$|\Delta v| = \sqrt{5.0^2 + 7.5^2}$$

$$= 9.0 \text{ m/s}$$

$$\theta = \tan^{-1}\left(\frac{5.0}{7.5}\right) = 33.7^\circ$$



The change in velocity is 9.0 m/s 33.7° west of south, or with bearing of

$$180 + 33.7 = 214^\circ$$



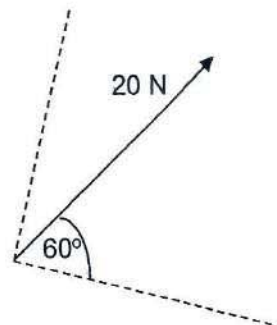
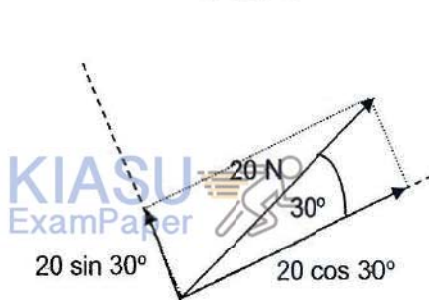
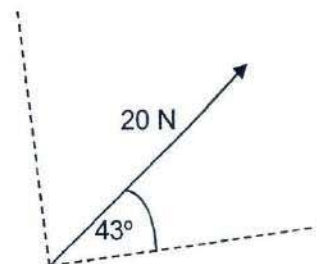
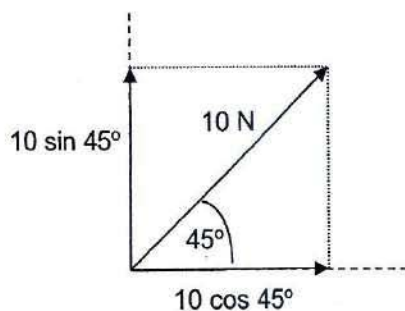
1.4.2 Resolution of Vectors

Since two vectors can be added to give a resultant vector, any vector can be broken up (or resolved) into two vectors or components. We will usually resolve a vector into two mutually-perpendicular components through the use of trigonometry and Pythagoras theorem. Mutually-perpendicular vectors are independent of each other.

Practice

For each vector below, draw the two perpendicular components in the direction given by the dotted lines and state their magnitudes in terms of the respective values and angles.

The examples on the left are done for you. Note that the dotted rectangle is an important working to account for the exact magnitudes of component vectors.



Each of the vectors above is resolved into two perpendicular components. A vector can be resolved into infinite pairs of perpendicular components. The choice of directions depends on the problem at hand.

Example 1.16

- (a) An object rests on the plane of an inclined slope as shown. The weight W acts vertically down. Draw components of the weight

- (i) parallel to the slope
(ii) perpendicular (normal) to the slope.

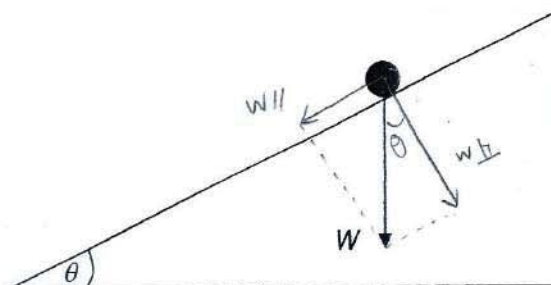
Watch detailed worked solution:



Label the magnitude of the two components in terms of W and θ .

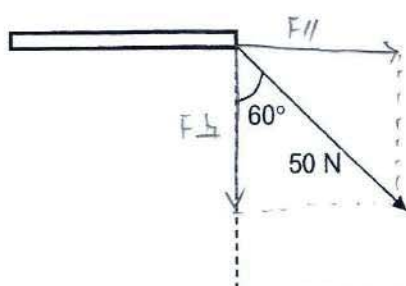
$$W_{\parallel} = W \sin \theta$$

$$W_{\perp} = W \cos \theta$$



- (b) A force of 50 N acts on a horizontal plank at angle of 60° to the vertical as shown. Draw components of this force (a) parallel to the plank (b) perpendicular to the plank.

Determine the magnitude of these components.



$$a) F_{\parallel} = 50 \sin 60^\circ = 43.3 \text{ N}$$

$$b) F_{\perp} = 50 \cos 60^\circ = 25 \text{ N}$$

Example 1.17

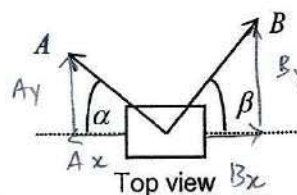
Two forces A and B act on an object at angles α and β as shown in the diagram. Determine the expression for the resultant force.

$$R_x = -A_x + B_x = -A \cos \alpha + B \cos \beta$$

$$R_y = A_y + B_y = A \sin \alpha + B \sin \beta$$

$$R = \sqrt{R_x^2 + R_y^2} = \sqrt{(-A \cos \alpha + B \cos \beta)^2 + (A \sin \alpha + B \sin \beta)^2}$$

$$\text{Angle with respect to the horizontal} = \tan^{-1} \left(\frac{A \sin \alpha + B \sin \beta}{-A \cos \alpha + B \cos \beta} \right)$$



KIASU
ExamPaper

Watch detailed worked solution:



Example 1.18

The diagram shows the top view of three motorboats crossing a river (left bank to right bank). All three move at the same speed but in different directions, and all experience the same current flowing down the river.

Construct resultant vectors showing the speed and direction of the boats.

- a. Which boat takes the shortest path to the opposite shore?

A

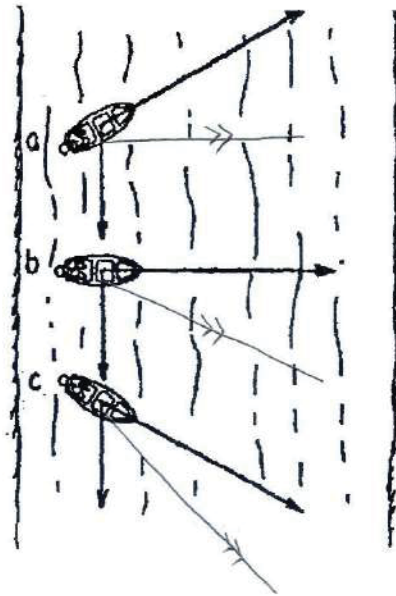
- b. Which boat reaches the opposite shore first?

$$t = \frac{d}{v}$$

B

- c. Which boat provides the fastest ride?

C



Watch detailed worked solution:

**Further reference:**

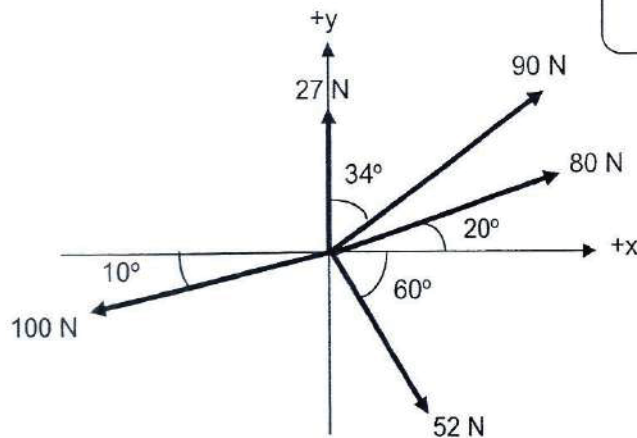
MIT OpenCourseWare Video - Vectors



Example 1.19

The 5 forces shown act on an object. Find the resultant force.

Watch detailed worked solution:

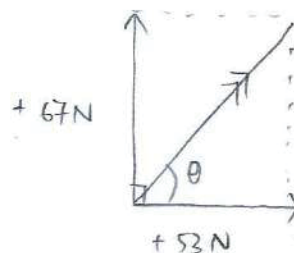


- Resolve the vectors into two mutually perpendicular components:

Vector/N	x-component /N (+ →)	y-component /N (+↑)
80	$80 \cos 20^\circ$	$80 \sin 20^\circ$
90	$90 \sin 34^\circ$	$90 \cos 34^\circ$
27	0	27
100	$-100 \cos 10^\circ$	$-100 \sin 10^\circ$
52	$52 \cos 60^\circ$	$-52 \sin 60^\circ$
Resultant	53	67

- Draw vector diagram to show the resultant vector.

- Compute the magnitude of the resultant vector. [85 N]



- Indicate the direction with an angle with reference to a certain direction on the diagram and calculate the angle. [52° anticlockwise from +ve x-direction]

$$R = \sqrt{67^2 + 53^2} = 85 \text{ N}$$

$$\theta = \tan^{-1} \frac{67}{53} = 52^\circ$$

The resultant force is 85 N, 52° anticlockwise from the +ve x-axis.

- Write a complete statement to specify the magnitude and direction of the resultant vector

Further Reading and References

Appendix I International System of Units (SI)

The seven SI base quantities and units of the International System of Units, universally abbreviated SI (from the French *Le Système International d'Unités*), established in 1960 by the 11th General Conference on Weights and Measures (CGPM, Conférence Générale des Poids et Mesures).

Base Quantity	SI Base Unit		
	Name	Symbol	Definition
length	metre	m	The metre is the length of the path travelled by light in vacuum during a time interval of $1/299\,792\,458$ of a second.
mass	kilogram	kg	The kilogram is the unit of mass; it is equal to the mass of the international prototype of the kilogram. (The prototype is made of platinum-iridium, is kept at the BIPM under conditions specified by the 1st CGPM in 1889.)
time	second	s	The second is the duration of $9\,192\,631\,770$ periods of the radiation corresponding to the transition between the two hyperfine levels of the ground state of the caesium 133 atom.
electric current	ampere	A	The ampere is that constant current which, if maintained in two straight parallel conductors of infinite length, of negligible circular cross-section, and placed 1 meter apart in vacuum, would produce between these conductors a force equal to 2×10^{-7} newton per meter of length.
thermodynamic temperature	kelvin	K	The kelvin, unit of thermodynamic temperature, is the fraction $1/273.16$ of the thermodynamic temperature of the triple point of water.
amount of substance	mole	mol	The mole is the amount of substance of a system which contains as many elementary entities as there are atoms in 0.012 kilogram of carbon 12.
luminous intensity	candela	cd	The candela is the luminous intensity, in a given direction, of a source that emits monochromatic radiation of frequency 540×10^{12} hertz and that has a radiant intensity in that direction of $1/683$ watt per steradian.

Adapted from the National Institute of Standards and Technology (NIST) Reference on Constants, Units and Uncertainty.

Appendix II Summary of Key Quantity, Symbols and Units

SUMMARY OF KEY QUANTITIES, SYMBOLS AND UNITS

The following list illustrates the symbols and units that will be used in question papers.

<i>Quantity</i>	<i>Usual symbols</i>	<i>Usual unit</i>
<i>Base Quantities</i>		
mass	m	kg
length	l	m
time	t	s
electric current	I	A
thermodynamic temperature	T	K
amount of substance	n	mol
<i>Other Quantities</i>		
distance	d	m
displacement	s, x	m
area	A	m ²
volume	V, v	m ³
density	ρ	kg m ⁻³
speed	u, v, w, c	m s ⁻¹
velocity	u, v, w, c	m s ⁻¹
acceleration	a	m s ⁻²
acceleration of free fall	g	m s ⁻²
force	F	N
weight	W	N
momentum	p	N s
work	w, W	J
energy	E, U, W	J
potential energy	E_p	J
kinetic energy	E_k	J
heating	Q	J
change of internal energy	ΔU	J
power	P	W
pressure	p	Pa
torque	T	N m
gravitational constant	G	N kg ⁻² m ²
gravitational field strength	g	N kg ⁻¹
gravitational potential	ϕ	J kg ⁻¹
angle	θ	°, rad
angular displacement	θ	°, rad
angular speed	ω	rad s ⁻¹
angular velocity	ω	rad s ⁻¹
period	T	s
frequency	f	Hz
angular frequency	ω	rad s ⁻¹
wavelength	λ	m
speed of electromagnetic waves	c	m s ⁻¹
electric charge	Q	C
elementary charge	e	C
electric potential	V	V
electric potential difference	V	V
electromotive force	E	V
resistance	R	Ω
resistivity	ρ	$\Omega \text{ m}$
electric field strength	E	N C ⁻¹ , V m ⁻¹
permittivity of free space	ϵ_0	F m ⁻¹
magnetic flux	Φ	Wb

<i>Quantity</i>	<i>Usual symbols</i>	<i>Usual unit</i>
magnetic flux density	B	T
permeability of free space	μ_0	H m^{-1}
force constant	k	N m^{-1}
Celsius temperature	θ	$^{\circ}\text{C}$
specific heat capacity	c	$\text{J K}^{-1} \text{kg}^{-1}$
molar gas constant	R	$\text{J K}^{-1} \text{mol}^{-1}$
Boltzmann constant	k	J K^{-1}
Avogadro constant	N_A	mol^{-1}
number	N, n, m	
number density (number per unit volume)	n	m^{-3}
Planck constant	h	J s
work function energy	ϕ	J
activity of radioactive source	A	Bq
decay constant	λ	s^{-1}
half-life	$t_{1/2}$	s
relative atomic mass	A_r	
relative molecular mass	M_r	
atomic mass	m_a	kg, u
electron mass	m_e	kg, u
neutron mass	m_n	kg, u
proton mass	m_p	kg, u
molar mass	M	kg
proton number	Z	
nucleon number	A	
neutron number	N	

Appendix III Homogeneous Equation

A physically correct equation must be homogeneous. However, a homogeneous equation need not be physically correct.

There are two basic reasons:

- (1) The value of the coefficient may be incorrect.

e.g. $E = 3mv^2$ where E = kinetic energy

The coefficient 3 is incorrect. The value should be $\frac{1}{2}$ instead.

- (2) Missing or extra terms that may have the same unit.

e.g. $E = \frac{1}{2}mv^2 + mgh$ where E = kinetic energy

There is an extra term mgh , which happens to have the same base unit as kinetic energy. This is an extra term.

Appendix IV SI Prefixes^f

Prefix	Symbol	Decimal	Power of ten	Order of magnitude
yocto-	y	0.000000000000000000000001	10^{-24}	-24
zepto-	z	0.00000000000000000000001	10^{-21}	-21
atto-	a	0.0000000000000000000001	10^{-18}	-18
femto-	f	0.000000000000000001	10^{-15}	-15
pico-	p	0.00000000000001	10^{-12}	-12
nano-	n	0.000000001	10^{-9}	-9
micro-	μ	0.000001	10^{-6}	-6
milli-	m	0.001	10^{-3}	-3
centi-	c	0.01	10^{-2}	-2
deci-	d	0.1	10^{-1}	-1
-	-	1	10^0	0
deca-	da	10	10^1	1
hecto-	h	100	10^2	2
kilo-	k	1,000	10^3	3
mega-	M	1,000,000	10^6	6
giga-	G	1,000,000,000	10^9	9
tera-	T	1,000,000,000,000	10^{12}	12
peta-	P	1,000,000,000,000,000	10^{15}	15
exa-	E	1,000,000,000,000,000,000	10^{18}	18
zetta-	Z	1,000,000,000,000,000,000,000	10^{21}	21
yotta-	Y	1,000,000,000,000,000,000,000,000	10^{24}	24

For use in Information Technology, prefixes for binary multiples have been adopted by the International Electrotechnical Commission (IEC).

Please refer to website for more information: <http://physics.nist.gov/cuu/Units/binary.html>



It appears that everyone has been wrong about the kilobyte so you can read for yourself.

^f http://en.wikipedia.org/wiki/Order_of_magnitude

Appendix V Rules of Significant Figures in Mathematical Operations

The following set of rules for **significant figures** provide us a means to get reliable results from data even without uncertainty values.

Rule 1: When multiplying or dividing, the number of s.f. in the answer should be the same as the term with the least numbers of s.f.

e.g. multiplying:

$$\begin{array}{r} 2.61 \\ \times 1.2 \\ \hline 3.132 \end{array}$$

← this has the smaller number of s.f. (2 s.f.)
→ rounds off to 3.1 (2 s. f.)

e.g. dividing: $2.61 \div 1.2 = 2.2$ (rounds off to 2 s.f.)

Practice: $2.3 \times 10.0 =$
 $6.8907 \div 1.23 =$

Rule 2: When adding or subtracting, decimal places (d.p.) in the answer are determined by the term with the least number of d.p., regardless of the s.f. of any one term.

e.g. adding:

$$\begin{array}{r} 26.46 \\ + 0.1234 \\ \hline 26.5834 \end{array}$$

← this has the least number of d.p. (2 d.p.)
→ rounds off to 26.58 (2 d.p.)

e.g. subtracting:

$$26.46 - 4.123 = 22.34 \text{ (to 2 d.p.)}$$

Practice: $5.399 - 5.39 =$
 $103.9 + 2.10 + 0.319 =$

More practices: <http://lectureonline.cl.msu.edu/~mmp/applist/sigfig/sig.htm> (requires Java)

Rule 3: For any logarithm, the number to the left of the decimal point is called the **characteristic, and the number to the right of the decimal point is called the **mantissa**.**

$$\text{e.g. } \lg 339 = \underline{2.530}$$

The mantissa has as many significant figures as the number which we took logarithms. The characteristic is not included when determining the number of significant figures.

$$\begin{aligned} \text{e.g. } \lg (5.43 \times 10^{10}) &= 10.735 \\ \lg (2.7 \times 10^{-8}) &= -7.57 \\ \ln (3.95 \times 10^6) &= 15.189 \end{aligned}$$

KIASU
ExamPaper 

Topic 2: Kinematics

H2 Physics C1 2016
H2 Physics Syllabus 9646

Topic 2: Kinematics

Kinematics	Learning Outcomes
	Students should be able to:
Rectilinear motion	(a) define and use displacement, speed, velocity and acceleration. (b) use graphical methods to represent distance, displacement, speed, velocity and acceleration. (c) identify and use the physical quantities from the gradients of displacement-time graphs and areas under and gradients of velocity-time graphs, including cases of non-uniform acceleration. (d) derive, from the definitions of velocity and acceleration, equations which represent uniformly accelerated motion in a straight line. (e) solve problems using equations which represent uniformly accelerated motion in a straight line, including the motion of bodies falling in a uniform gravitational field without air resistance. (f) describe qualitatively the motion of bodies falling in a uniform gravitational field with air resistance.
Projectile motion	(g) describe and explain motion due to a uniform velocity in one direction and a uniform acceleration in a perpendicular direction.

2.0 Introduction

Kinematics (from Greek κινεῖν, *kinein*, to move) is the branch of mechanics that describes the motion of objects without consideration of the causes leading to the motion. The study of why objects move falls under the branch of *Dynamics*.

2.1 Terminology

2.1.1 Distance vs. Displacement

Distance, x

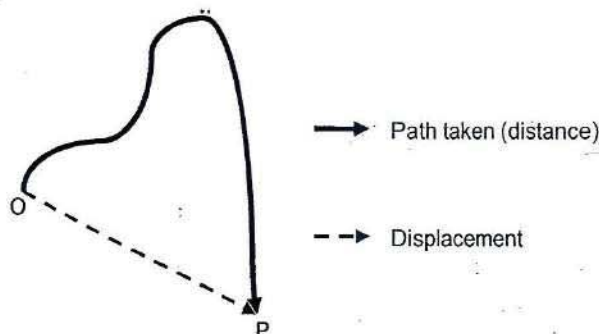
Distance is the total length travelled by a moving object irrespective of the direction of motion.

Displacement, s

The **displacement** of an object from a reference point O is the **linear distance** and **direction** of the object from O.

Note that distance is a **scalar** quantity whereas displacement is a **vector**¹ quantity.

For illustration, in the diagram below, an object travels from point O to point P along the path represented by the solid line. The distance travelled is represented by the length of the solid line. The displacement is represented by the **length and direction** of the dashed line.



2.1.2 Speed vs. Velocity

Speed, v

The speed of an object is defined as the **rate of change of distance** travelled by an object with respect to **time**.

Velocity, v

The velocity of an object is its **rate of change of displacement** with respect to **time**.

Note that speed is a **scalar** quantity whereas velocity is a **vector** quantity.

¹ Generally, scalar quantities in textbooks are italicised and non-bold, e.g. x , while vector quantities in textbooks are either represented by an arrow above the quantity, a twiddle ($\sim$) below the quantity or italicised and bold, e.g. $\vec{A}$, $\underline{a}$, or simply A .



In past examinations, students were given zero marks for defining speed as distance travelled in one second. Quantities should always be defined in terms of other quantities and units should always be defined in terms of other units.

Mathematically, average and instantaneous speedⁱⁱ and velocity can be calculated as follow:

$$\begin{aligned}\text{Average speed, } \langle v \rangle &= \frac{\text{Total distance travelled}}{\text{Total time taken}} \\ &= \frac{\Delta x}{\Delta t}\end{aligned}$$

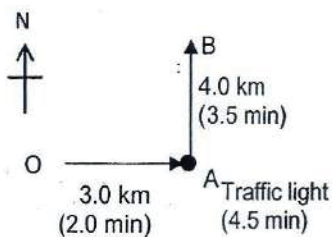
$$\text{Instantaneous speed, } v = \frac{dx}{dt}$$

$$\begin{aligned}\text{Average velocity, } \langle v \rangle &= \frac{\text{Total change in displacement}}{\text{Total time taken}} \\ &= \frac{\Delta s}{\Delta t}\end{aligned}$$

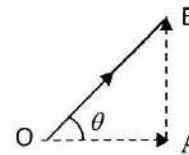
$$\text{Instantaneous velocity, } v = \frac{ds}{dt}$$

Example 1

A car travels 3.0 km due east from a point O to a point A for 2.0 minutes where it stops at a traffic light at A for 4.5 minutes. It then continues 4.0 km due north of A to point B for another 3.5 minutes. Find the average speed and average velocity of the car.



$$\begin{aligned}\text{Average speed} &= \frac{\text{total distance travelled}}{\text{time taken}} \\ &= \frac{3.0 + 4.0}{2.0 + 4.5 + 3.5} \\ &= 0.70 \text{ km min}^{-1} \\ &= 12 \text{ m s}^{-1}\end{aligned}$$



$$\begin{aligned}\text{Average velocity} &= \frac{\text{total change in displacement OB}}{\text{time taken}} \\ &= \frac{\sqrt{3.0^2 + 4.0^2}}{2.0 + 4.5 + 3.5} \\ &= 0.50 \text{ km min}^{-1} = 8.3 \text{ m s}^{-1} \\ \tan \theta &= \frac{4}{3} \\ \theta &= 53^\circ\end{aligned}$$

Note that as velocity is a vector quantity, it is necessary to find the direction for average velocity.

ⁱⁱ It is rare that the speed of an object is constant. When we refer to the speed of the object, we are usually referring to the **instantaneous speed**. Similarly when we refer to **velocity** and **acceleration**, it is the **instantaneous velocity** and **instantaneous acceleration** that is referred to, unless stated otherwise.

Example 2

A particle moves with an initial velocity of 10 m s^{-1} in the x -direction for 5.0 s . It then changes its velocity to 30 m s^{-1} at an angle of 60° to the horizontal and moves for a further 2.0 s . (Figure 2.1)

- Calculate the change in speed and velocity of the particle in this interval.
- Find the average speed and average velocity of the particle.

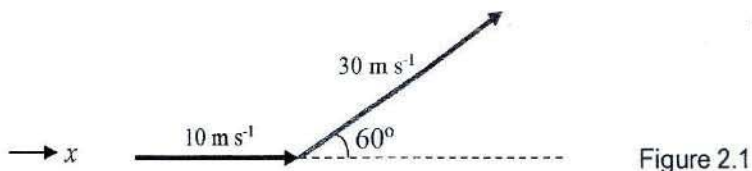


Figure 2.1

Solution

- Change in speed = final speed – initial speed
 $= (30 - 10) \text{ m s}^{-1}$
 $= 20 \text{ m s}^{-1}$

$$\text{Change in velocity} = v_{\text{final}} - v_{\text{initial}} = v_{\text{final}} + (-v_{\text{initial}})$$

Magnitude of the change in velocity

Using cosine rule,

$$\Delta v = \sqrt{10^2 + 30^2 - 2(10)(30) \cos 60^\circ}$$

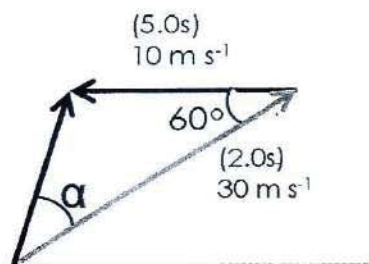
$$= 26.5 \text{ m s}^{-1}$$

Direction of the change in velocity

Using sine rule,

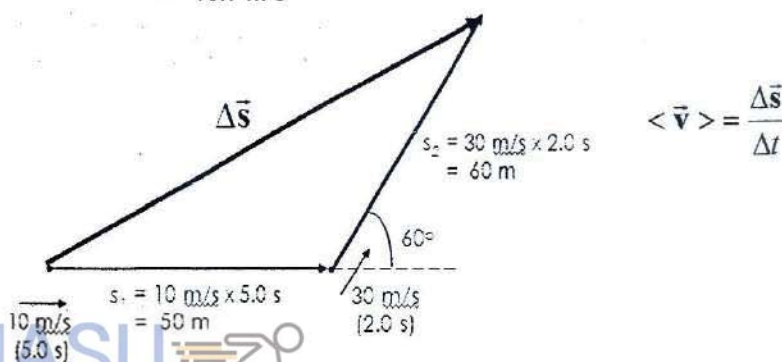
$$\frac{\sin \alpha}{10} = \frac{\sin 60^\circ}{26.5}$$

$$\alpha = 19.1^\circ$$



The change in the velocity is 26.5 m s^{-1} at an angle of 79.1° anticlockwise from the horizontal

- Average speed = $\frac{\text{total distance travelled}}{\text{total time taken}}$
 $= \frac{(10 \times 5.0) + (30 \times 2.0)}{(5.0 + 2.0)}$
 $= 15.7 \text{ m s}^{-1}$



$$\text{Using cosine rule, } \Delta s = \sqrt{60^2 + 50^2 - 2(60)(50) \cos 120^\circ} = 95.4 \text{ m}$$

$$\text{Magnitude of the average velocity, } |\bar{v}| = 95.4 / 7.0 = 13.6 \text{ m s}^{-1}$$

$$\text{Using sine rule, } \beta = 33.0^\circ$$

The average velocity is 13.6 m s^{-1} at 33.0° to the initial velocity.

2.1.3 Acceleration

Acceleration, a

The **acceleration** of an object is its **rate of change of velocity** with respect to time.

Mathematically, average and instantaneous acceleration can be calculated as follow:

$$\begin{aligned}\text{Average acceleration, } \langle a \rangle &= \frac{\text{Total change in velocity}}{\text{Total time taken}} = \frac{\Delta v}{\Delta t} \\ &= \frac{v_{\text{final}} - v_{\text{initial}}}{\Delta t}\end{aligned}$$

$$\text{Instantaneous acceleration, } a = \frac{dv}{dt}$$

- S. I. units of acceleration: m s^{-2}
- Acceleration is a vector quantity. The direction of acceleration is in the same direction as the vector representing the **change** in velocityⁱⁱⁱ.

TIPS Sign Convention

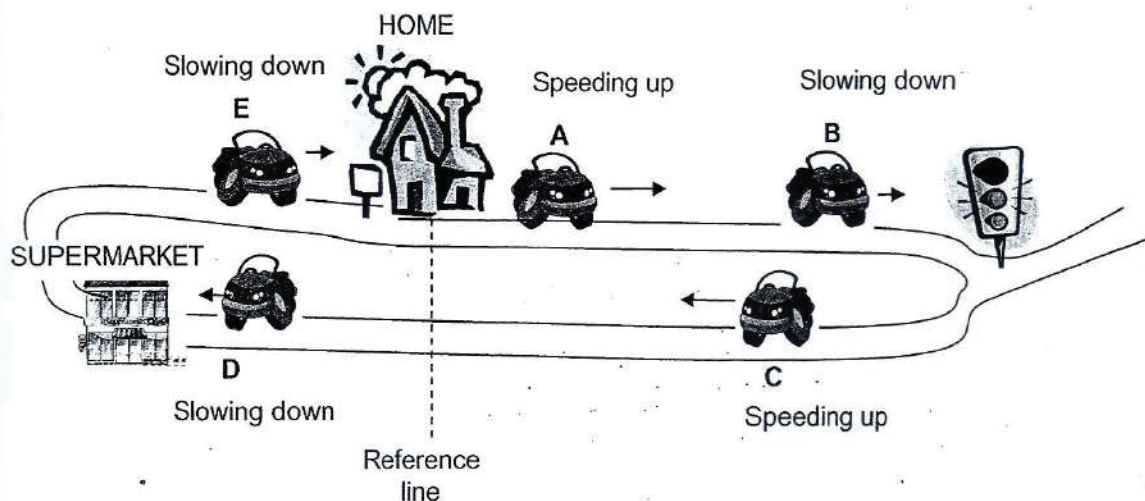
1. For an object moving in one-dimension (rectilinear), there are only two possible directions in which the object can move. One direction is opposite the other. Hence, for convenience, we define one direction as positive and the other negative.
2. For vertical motion, we usually define upward direction as positive*: $\uparrow +ve$
3. For horizontal motion, we usually take direction to the right as positive*: $\rightarrow +ve$

*But this depends on the problem given, if it is more convenient to define downwards as positive or leftwards as positive, you may do so. In any case, **you should always state clearly the convention you have adopted.**

ⁱⁱⁱ You will learn in Dynamics that the direction of any object's acceleration is determined by the direction of the resultant force acting on the object (Newton's 2nd Law of Motion).

Example 3

John drove a car from his home towards the traffic junction. As the traffic light turned red, he stopped and waited for his turn. He then made a U-turn and drove to a supermarket to buy some groceries. John drove back to his home after making the stop at the supermarket. The journey is illustrated in the figure below, where points A, B, C, D and E are located along his journey.



The reference line passes through John's home, and displacement to the right of it is defined as positive. Indicate the sign for John's displacement, velocity and acceleration at A, B, C, D and E.

diff when slowing down

	Displacement	Velocity	Acceleration
A (speeding up)	+	+	+
B (slowing down)	+	+	-
C (speeding up)	+	-	-
D (slowing down)	-	-	+
E (slowing down)	-	+	-

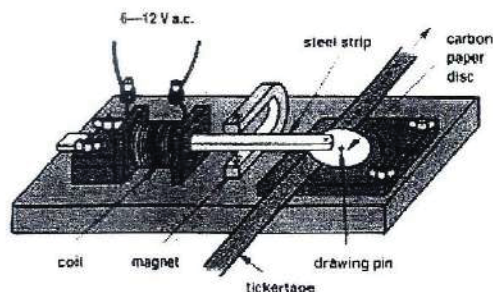
make velocity more "-ve"



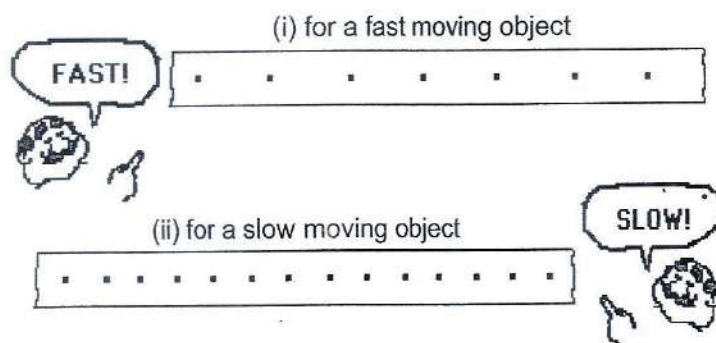
2.2 Describing Motion using Diagrams

2.2.1 Ticker tape diagrams

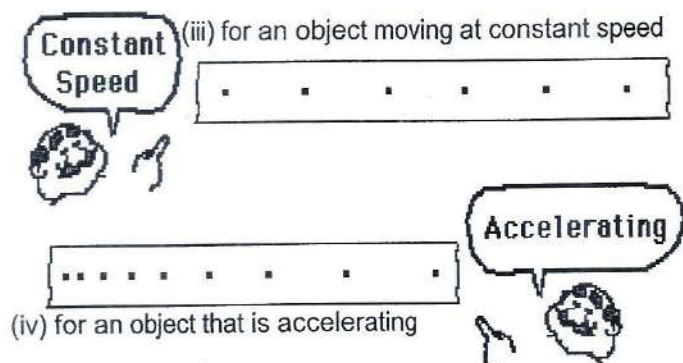
A common way of analyzing the motion of objects in physics labs is to perform a ticker tape analysis. A long tape is attached to a moving object and threaded through a device that places a tick upon the tape at regular intervals of time – say every 0.1 second. As the object moves, it drags the tape through the "ticker," thus leaving a trail of dots. The trail of dots provides a history of the object's motion and is therefore a representation of the object's motion.



The **distance between dots** on a ticker tape represents the object's position change during that time interval.

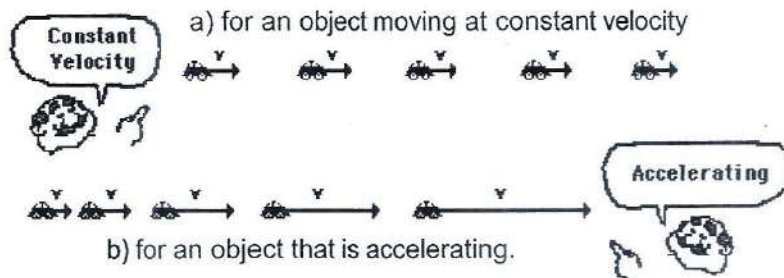
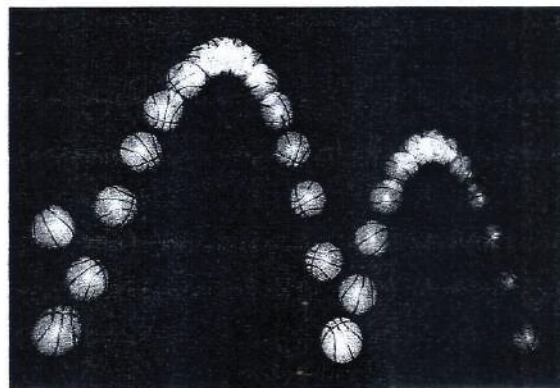


The analysis of a ticker tape diagram will also reveal if the object is moving with a constant velocity or with a changing velocity (accelerating).



2.2.2 Vector diagrams

Vector diagrams use vector arrows to depict the direction and relative magnitude of a vector quantity. Vector diagrams can be used to describe the velocity of a moving object during its motion. For example, the velocity of a car moving down the road could be represented by a vector diagram. In a vector diagram, the **magnitude** of the vector is represented by the length of the **vector** arrow.

2.2.3 Recording motion using stroboscope^{iv}

Multi-flash photo of a bouncing basketball

Multi-flash photographs or stroboscopic photographs are photographs of a subject illuminated with a flashing light, taken by a camera with an open shutter. By keeping the subject in the dark and then illuminating the subject only at certain times, the resultant image will show a series of different positions of the subject.

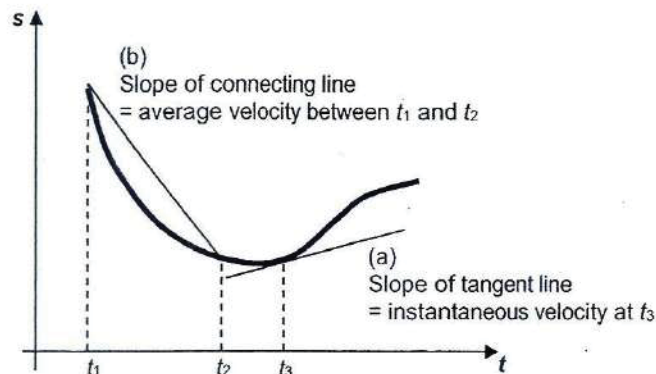
Today, by using digital video camera and digital image compositing, a multi-flash photograph can be easily re-constructed from several video frames. The video camera enables us to record motion conveniently, even at high speed. A single video camera captures 2-dimensional motion on a plane parallel to the image plane of the camera. To record complex 3-dimensional motion, many video cameras are used. This technique is useful for sports biomechanical analysis and character analysis in animated film production.



^{iv} <http://edgerton-digital-collections.org/techniques/multiflash>

2.3 Describing Motion using Graphs

2.3.1 Displacement vs. Time Graph (s - t graph)

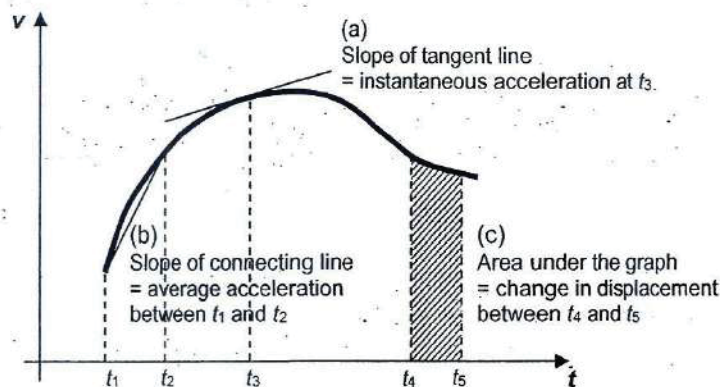


On an s - t graph,

The **instantaneous velocity** is the **slope** (or **gradient**) of the **tangent line** at a given **instant** of time.

- Average velocity is the **slope** of the **straight line connecting two points** corresponding to a given time **interval**.
- If an object moves with **constant** velocity, its s - t graph will be a **straight** line graph, and the instantaneous velocity at any instant is equal to the average velocity over the entire time interval.

2.3.2 Velocity vs. Time Graph (v - t graph)



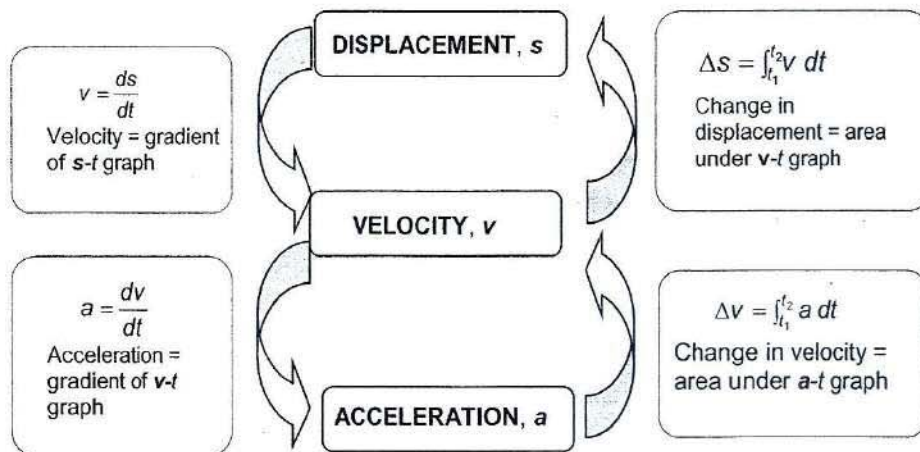
On an v - t graph,

The **instantaneous acceleration** is the **slope** (or **gradient**) of the **tangent line** at a given instant of time.

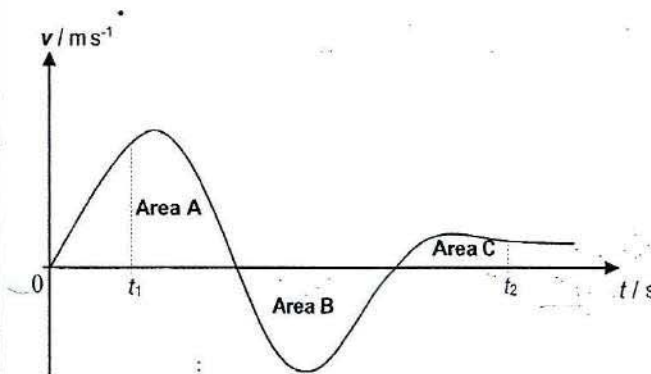
- Average acceleration is the **slope** of the **straight line connecting two points** corresponding to a given time **interval**.
- If an object moves with **constant** acceleration, its v - t graph will be a **straight** line graph, and the instantaneous acceleration at any instant is equal to the average acceleration over the entire time interval.

The **area under a v - t graph** for a time interval of Δt gives the **change in displacement** of the object in that time interval.

2.3.3 Summary of relations among s-t, v-t and a-t graphs

**Example 4**

What is the expression of the change in displacement, Δs , of the object from t_1 to t_2 ?
If you need to find distance travelled by the object in the same time interval, what would be your answer?



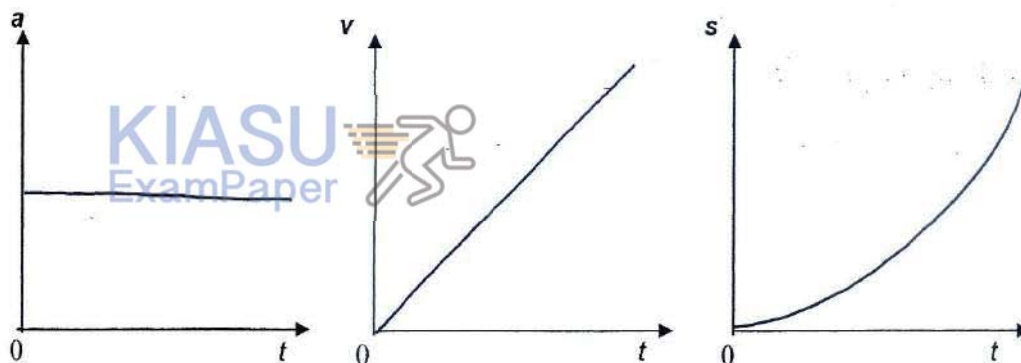
$$\Delta s = \text{Area A} + (-\text{Area B}) + \text{Area C}$$

Distance travelled by the object in the same time interval,

$$\Delta x = \text{Area A} + \text{Area B} + \text{Area C}$$

Example 5

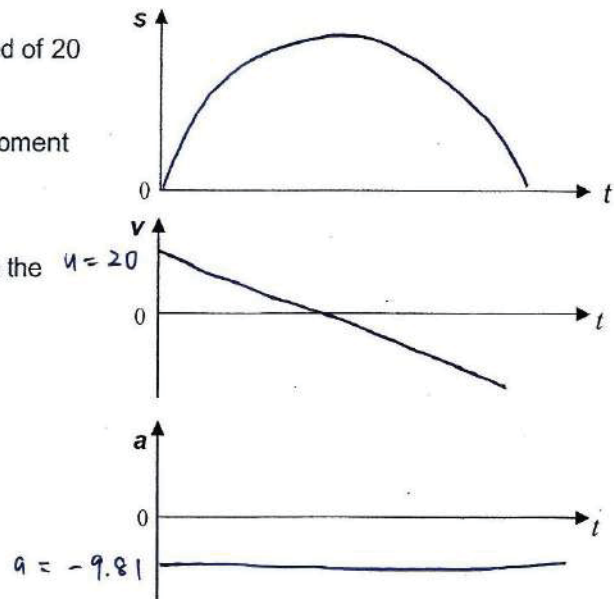
An object is at rest at the origin. If the object undergoes constant acceleration, sketch its a-t, v-t and s-t graphs.



Example 6

A student throws a ball vertically upwards at a speed of 20 m s^{-1} from level A. Neglect air resistance.

- Sketch the $s-t$, $v-t$ and $a-t$ graphs from the moment the ball is thrown to the moment it returns to the student's hand.
- Calculate the time it takes for the ball to return to the student's hand.



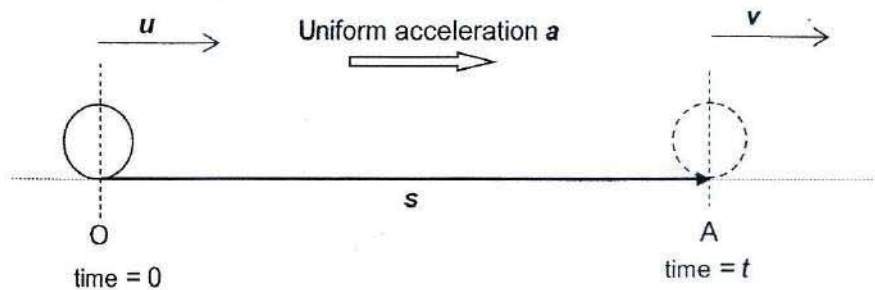
$$\begin{aligned}
 \text{(ii). gradient} &= -9.81 = (0 - 20) / t \\
 t &= 2.04 \text{ s} \\
 \text{time taken} &= 2 \times 2.04 \\
 &= 4.08 \text{ s}
 \end{aligned}$$

2.4 Describing Motion using Equations (for uniform acceleration)

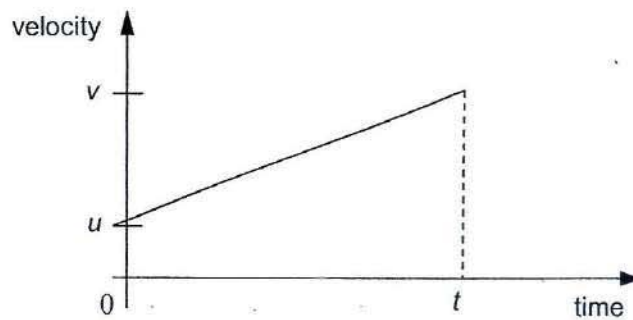
2.4.1 Derivation of the Equations of Motion (for uniform acceleration)

In this section, we shall derive the equations of motion for an object moving in a straight line with **uniform acceleration**.

Suppose at time = 0, the object is at point O with an initial velocity of u . The object moves at uniform acceleration of a until time = t , and has travelled through a displacement of s to arrive at point A with the final velocity of v .



Let us look at its velocity-time graph:



Note: acceleration of object, a is given by the gradient of its $v-t$ graph,

$$a = \frac{v - u}{t}$$

$$v = u + at$$

(1)

Note: The displacement of object from point O, s , is given by the area under its $v-t$ graph from time = 0 to time = t ,

$$s = \frac{1}{2}(u + v)t$$

(2)

Substitute Eqn (1) into (2), we have

$$s = \frac{1}{2}[u + (u + at)]t$$

$$\Rightarrow s = ut + \frac{1}{2}at^2$$

(3)

From Eqn (1), we have $t = \frac{v - u}{a}$ and substitute into Eqn (2), we have

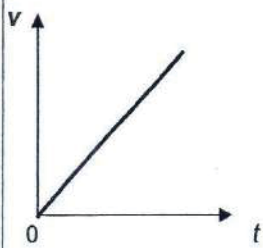
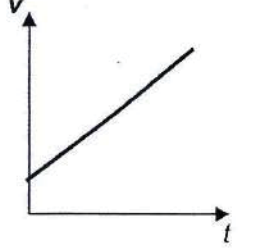
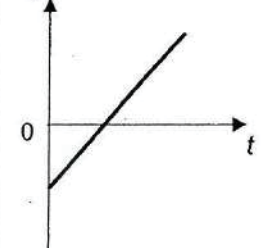
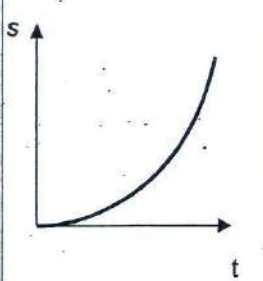
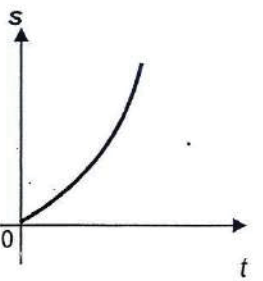
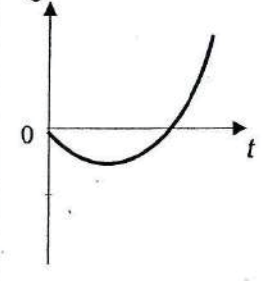
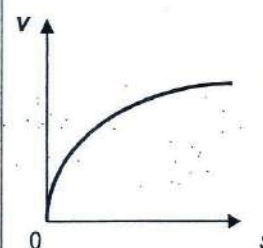
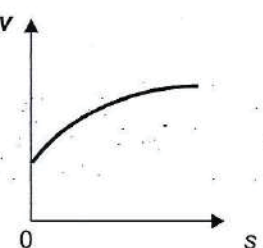
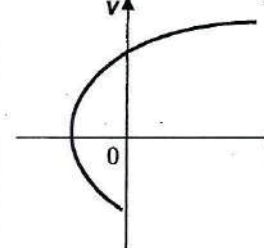
$$s = \frac{1}{2}(u + v)\left(\frac{v - u}{a}\right)$$

$\Rightarrow$

$$v^2 = u^2 + 2as$$

(4)

In summary, the equations of motion are tabulated below.

Equations of Motion	Graphical Representation		
	Case 1 (initial velocity is zero, acceleration positive)	Case 2 (initial velocity is positive, acceleration positive)	Case 3 (initial velocity is negative, acceleration positive)
$v = u + at$ Use this equation to relate velocity to time. v - t graphs are straight line graphs.			
$s = ut + \frac{1}{2}at^2$ Use this equation to relate displacement to time. s - t graphs are quadratic curves (The shape is called a parabola.)			
$v^2 = u^2 + 2as$ Use this equation to relate velocity to displacement. v - s graphs follow the shape of a square-root graph.			

Pause and Think: Why do you think we did not include the graph for $s = \frac{1}{2}(u+v)t$ here?

Note:

- The equation ($s = ut + \frac{1}{2}at^2$) sets the reference point $s = 0$ at $t = 0$.
- Equations of motion can only be used if **the acceleration is constant**.
- The equations are **vector equations**. Hence the motion in each dimension is independent of the other direction. Therefore, variables also have to be analysed along the same dimension. For example, if we choose to analyse the motion along the x-direction, then we can only use quantities like u_x , v_x and a_x . Acceleration along the y-direction a_y will not affect motion along the x-direction.

KIASU
ExamPaper

Example 7: J82/P1/1

A motorist traveling at 13 m s^{-1} approaches traffic lights, which turn red when he is 25 m away from the stop line. His reaction time is 0.70 s. If he brakes fully such that the car slows down at a rate of 4.5 m s^{-2} ,

- how far from the stop line will he stop?
- on which side of the stop line will he come to a rest?

a) Dist travelled during the reaction time duration

$$= \text{speed} \times \text{time} = 13 \times 0.70 = 9.1 \text{ m}$$

Dist. travelled during braking

$$(\rightarrow) s = \frac{v^2 - u^2}{2a} = \frac{0 - 13^2}{2(-4.5)} = 18.8 \text{ m}$$

$$\text{Total dist} = 18.8 + 9.1 = 27.9 \text{ m}$$

b) She will stop $(27.9 - 25.0) = 2.9 \text{ m}$ after the stop line.

2.5 Free Falling Bodies

The simplest example of motion with constant acceleration is a free falling body.

A free falling object is any object moving only under the influence of gravity, i.e. the effects of resistive forces (i.e. drag forces) and upthrust are negligible. For free falling objects, whether they are initially thrown upward, downward or released from rest, they will accelerate downwards at a rate of 9.81 m s^{-2} ,^v regardless of the objects' weight, size and shape. Let us illustrate this below.

Consider a mass m under free falling. Since air resistance is negligible, the only force acting on the mass is its own weight. Applying Newton's 2nd Law,

$$F_{\text{net}} = ma$$

$$mg = ma$$

$$a = g$$

In essence, even though a larger mass experiences a larger gravitational force, its larger inertia also makes it harder to accelerate. These two effects cancel out cleanly. This curious coincidence causes a feather to fall at the same rate as a lead ball.

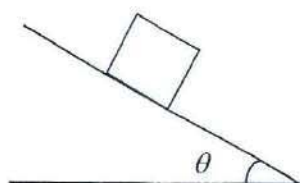
^v You will learn in the topic of Gravitation that the value of g does vary with location on earth. The average value of 9.81 m s^{-2} is conveniently chosen for general calculations.

Example 8: Are these Objects in Free-Fall?

The following are 2 cases of objects in motion.

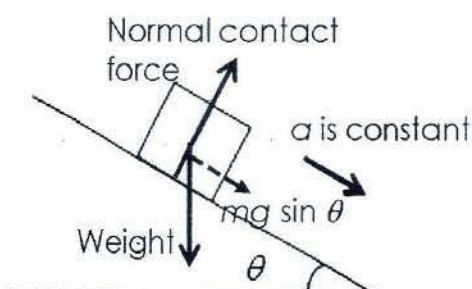
- Determine if each of the objects is in free fall.
- For each case, if the object is not in free fall, identify the physical quantities that affect the acceleration.
- Hence, explain if the equations of motion can be applied to each scenario.

Case 1: A box sliding down a smooth incline. Assume that air resistance is negligible.



Solution:

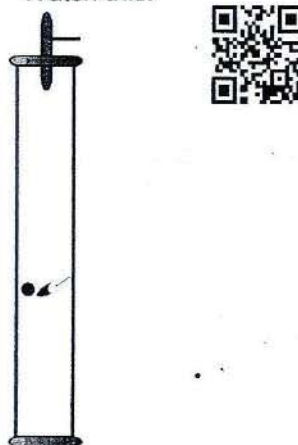
Considering the forces on the box,



- Not in free fall
- Net force: Component of weight of box parallel to the slope
Acceleration = $g \sin \theta$
Factors affecting acceleration : g and θ
- Since normal contact force and weight are constant, the resultant force acting on the box is constant. Hence, acceleration is constant and equation of motion can be applied in case 1.

Case 2: A feather falling in vacuum.

Watch this:



Solution:

Considering the force on the feather,



- In free fall
- Net force: Only weight of feather, W
($W = m_{\text{feather}} g$)
- $a = g$,
Hence, acceleration is constant and equation of motion can be applied in case 1.

For both cases, can you visualise the acceleration-time and velocity-time graphs?

2.6 Viscous Drag Force

In fluid dynamics, drag (sometimes called air resistance or fluid resistance) refers to forces that oppose the relative motion of an object through a fluid (a liquid or gas). When a body moves through a fluid, it has to push the surrounding fluid molecules away in order to move. Hence it exerts a force on the fluid molecules and by Newton's Third Law, the molecules exert an equal and opposite force back on the body. Hence there is resistance to motion.

It follows that drag force depends on the **velocity** of the body, *i.e.* the faster the motion of the body through the fluid, the larger the resistance. It also means that if the body is at rest relative to the fluid, there is no resistance.

Other factors affecting the magnitude of drag force include the **shape** and **dimension** of the body (a more streamline shape allows a body to experience less drag as it pushes through the fluid) and the type of fluid (**viscosity** of the fluid).

When there is no turbulent flow (e.g. small and streamline objects traveling at low speeds) the drag force is proportional to the speed of the object. The magnitude of drag force can be stated empirically as follows:

$$F_D = kv$$

F_D : magnitude of drag force

k : a constant that depends on the dimensions of the body and the type of fluid

v : the velocity of the body relative to the fluid

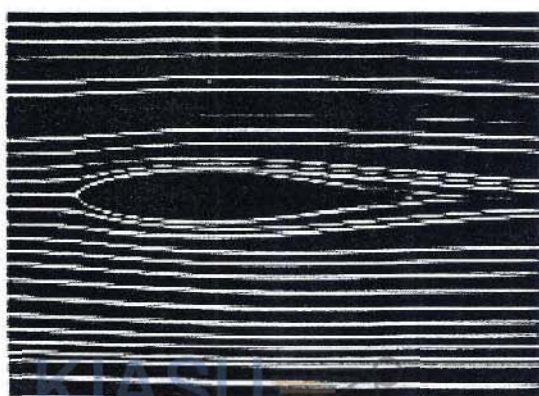
If the fluid flow is turbulent (e.g. objects moving at high speeds through air, such as aircraft, skydiver and baseball), the drag force is approximately proportional to the square of the speed. The drag force is then given by:

$$F_D = kv^2$$

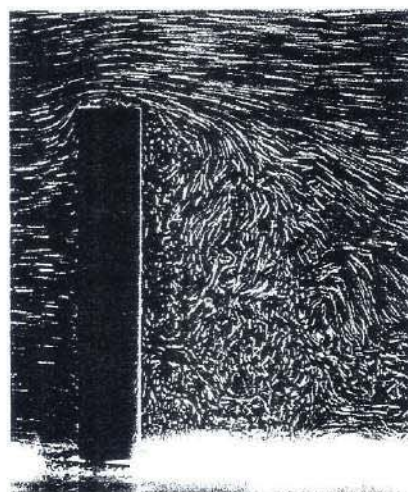
F_D : magnitude of drag force

k : a constant that depends on the dimensions of the body and the type of fluid

v : the velocity of the body relative to the fluid



Laminar flow over a streamline body



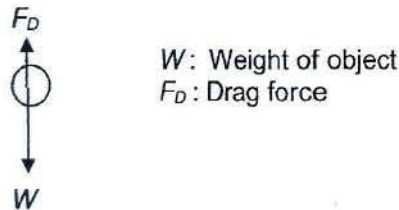
Turbulent flow results in higher drag

2.7 Non Free Falling Bodies

In practice, free fall does not happen because of resistive forces. The effects on a diver or bungee jumper may not be significant, but this is clearly not the case for a parachutist where air resistance is (deliberately) used to limit the maximum downward velocity of the parachutist.

2.7.1 Velocity-time graph of object falling through air

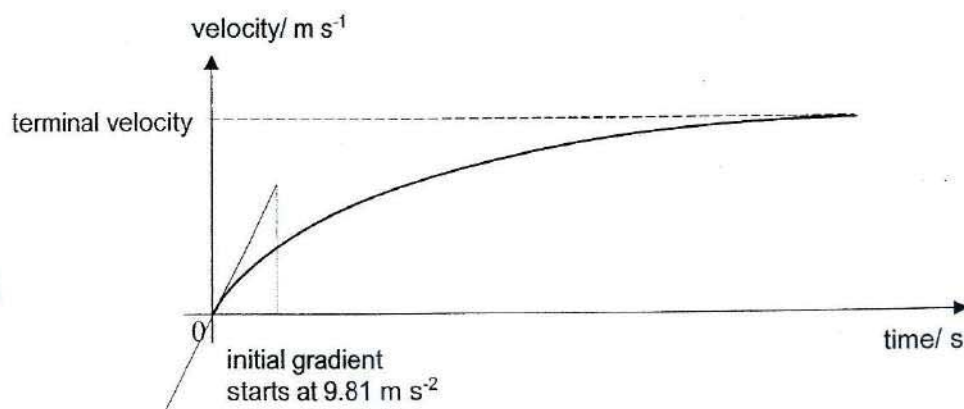
Let us now consider the motion of an object, dropped from rest, falling through air. If air resistance is not negligible, the forces acting on the object are:



For ease of analysis, break the motion up into stages as seen below. Let us assume that drag force is proportional to speed.

Time	Forces acting on the body	Motion of the object
$t = 0 \text{ s}$	$v = 0$ $W = mg$ a	- Object is initially at rest. - The only force acting on the object is the gravitational force (weight, $W = mg$). By Newton's 2nd Law, (1) $\Sigma F = ma \Rightarrow mg = ma \Rightarrow a = g$ - The body begins to fall with an acceleration of $a = g$.
$t = t_1$	$F_D = kv$ $W = mg$ a $v > 0$	- As the body gains velocity, air resistance which is dependent on the velocity also increases. By Newton's 2nd Law (1) $\Sigma F = ma \Rightarrow mg - kv = ma \Rightarrow a = g - \frac{kv}{m} < g$ - As F_D increases, the net force decreases and hence the acceleration decreases and falls below g. - The object is nevertheless still accelerating and hence the speed still increases (just at a decreasing rate), so the air resistance continues to increase.
$t = t_2$	$F_D = kv$ $W = mg$ $a = 0$ $v = v_{\text{terminal}}$ $v > 0$	- Eventually the air resistance increases to a point where it balances the weight of the body. i.e. $W = F_D$ By Newton's 2nd Law (1) $\Sigma F = ma \Rightarrow mg - kv = ma = 0 \Rightarrow a = 0$ - Resultant force acting on the body is zero. Hence acceleration of the body is zero. - The object then travels at a constant velocity called the terminal velocity.

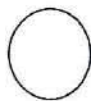
The following graph shows the non-linear velocity change with time for a falling object experiencing air resistance.



2.7.2 Terminal Velocity

We can now explain why heavier objects fall at a faster rate than light ones in the presence of air resistance. For simplicity, let us assume air resistance is proportional to speed v , i.e. $F_D = kv$.

Consider the free body diagram of an object falling at terminal velocity,



At terminal velocity, the acceleration is zero and so the net force is zero. Hence,

$$F_D = W$$

Assuming air resistance is proportional to speed, therefore $F_D = kv_t$

$$kv_t = mg$$

$$v_t = \frac{mg}{k}$$

Based on this equation, we can now conclude:

1. Between two objects of the same dimensions but different densities, the denser one reaches a higher terminal velocity because it has a larger m but the same k .
2. Between two objects of the same mass but different shapes, the more streamline one reaches a higher terminal velocity because it has a smaller k but the same m .

Example 9

A steel ball bearing is released from a height above a deep pool of oil. Taking the drag force between the ball bearing and oil to be proportional to the relative speed of the ball bearing in oil, sketch the graph to show the variation of the velocity v of the ball bearing against time t . Indicate clearly on your diagram, t_1 , the time when the ball bearing hits the oil.

Solution

Assume that upthrust is negligible.

Stage 1: Released from rest ($t = 0$ to t_1)

Before the ball bearing enters into the oil,

- Only the weight acts on it and drag force (due to air) is very small compared to weight.

By Newton's 2nd Law,

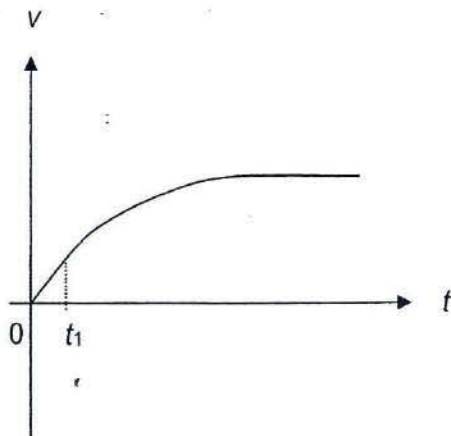
$$(\downarrow) \quad \Sigma F = ma$$

$$\Rightarrow mg = ma$$

$$\Rightarrow a = g$$

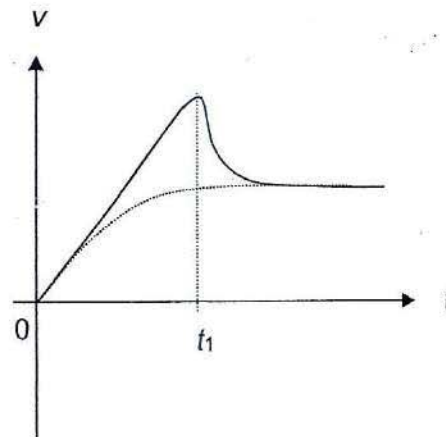
Stage 2 : Enters the oil

- Enters with some initial velocity
- Drag force due to oil resistance is potentially large compared to the weight, leading to two possible cases:
Case A: Ball bearing is released at/just above the surface of the oil, entering the oil at low velocity.
Case B: Ball bearing is released high above the surface of the oil, entering the oil at high velocity.



Case A: $F_{\text{drag}} < mg$

- Resultant force is downwards
- The ball bearing accelerates but at a value less than g .
- As v increases, F_{drag} increases until $F_{\text{drag}} = mg$
- Net force = 0
- Terminal velocity is reached



Case B: $F_{\text{drag}} > mg$

- Resultant force is upwards
- The ball bearing decelerates (slows down), v decreases
- As v decreases, F_{drag} decreases until $F_{\text{drag}} = mg$
- Net force = 0
- Terminal velocity is reached

2.8 Two-Dimensional Kinematics

In this part of the chapter, we deal with the kinematics of a particle moving in a plane, or two-dimensional motion. Some common examples of motion in a plane are the motion of projectiles, motion of satellites and the motion of charged particles in an electric field. For the beginning we shall deal with the special case of motion of **an object with a uniform velocity in one direction and a uniform acceleration in a perpendicular direction**.

2.8.1 Projectile Motion

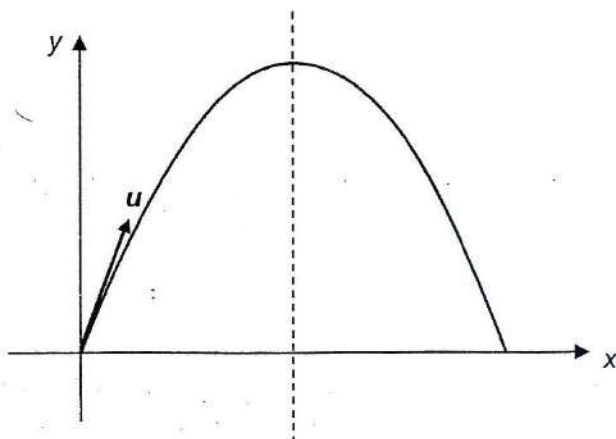
Anyone who has observed a soccer ball in motion (or, for that matter, any object thrown into air) has observed *projectile motion*.

Projectile motion refers to the motion of an object launched in the air in the presence of a gravitational field. In this section, we are interested in the motion of an object projected with an initial velocity at an angle to the horizontal. To analyze this type of motion, we will start with an idealized model. The assumptions made in our idealized model are:

- (1) the effect of air resistance is negligible.
- (2) the acceleration due to gravity g is constant^{vi}, acts downwards and has a magnitude equal to $g = 9.81 \text{ m s}^{-2}$.

With the above assumptions, we find that the path of a projectile, which we call its **trajectory**, is always a **parabola**^{vii}.

Key Characteristics of a Parabolic Path:



- The path of the projectile is symmetrical about the highest point.
- The speed of the projectile is symmetrical about the highest point.
- The time it takes for the object to go from $y = 0$ to the highest point
is equal to
 the time it takes for the object to go from the highest point to $y = 0$.

It was Galileo who first accurately described projectile motion. He showed that it could be understood by analyzing the horizontal and vertical components of the motion separately. He found that:

The horizontal and vertical motions of the projectile are independent of each other.

KIASU
ExamPaper

^{vi} We shall restrict ourselves to objects whose distance travelled and maximum height above the Earth are small compared to the Earth's radius (6400 km) so that g can be considered constant.

^{vii} The proof that trajectory of a projectile is parabolic will be shown in Example 13.

To examine this idea further, let us examine the motion of 2 balls:

- (1) Ball 1 is a ball released from rest and allowed to fall vertically through a hole on the table, and
- (2) Ball 2 is allowed to roll off the end of a table with an initial velocity u_x in the horizontal (x) direction.

Both the balls leave the table at the same time. The stroboscopic photograph below depicts the motion of both balls. Let us examine the vertical and horizontal motion separately.

○ **Vertical Motion:**

Consider the motion of both balls in the vertical direction.

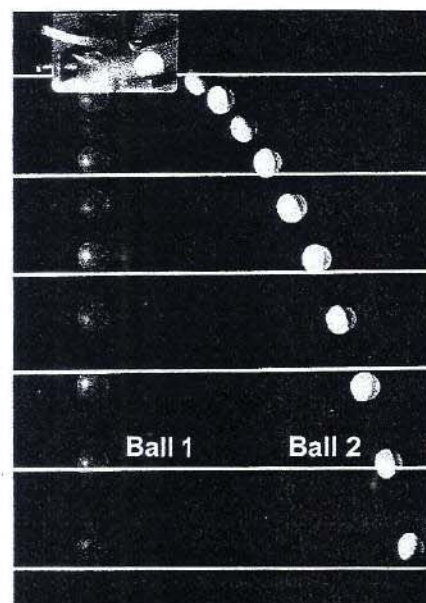
Note,

- The **vertical displacement of Ball 2 matches the vertical displacement of Ball 1.**
- The **vertical displacement and vertical velocity of Ball 2 are only affected by the vertical acceleration g (as in Ball 1).**

○ **Horizontal Motion:**

Consider the motion of Ball 2. Note that

- The **horizontal displacements made by the ball at equal intervals of time are approximately the same.**
- The **horizontal velocity remains constant with time.**
- The **horizontal acceleration is zero.**



Consistent with our assumption that only the weight acts on Ball 2 and no other forces act in the horizontal direction. Hence, horizontal acceleration is zero.

- The **vertical acceleration hence has no effect on the horizontal velocity and horizontal displacements.**

Hence, we see that horizontal and vertical motions are **independent** of each other and projectile motion can be simplified by analysing the vertical and horizontal motions separately. Another result of this analysis, which Galileo himself predicted, is that *an object projected horizontally will reach the ground in the same time as an object dropped vertically*. This is because the vertical motions are the same in both cases.

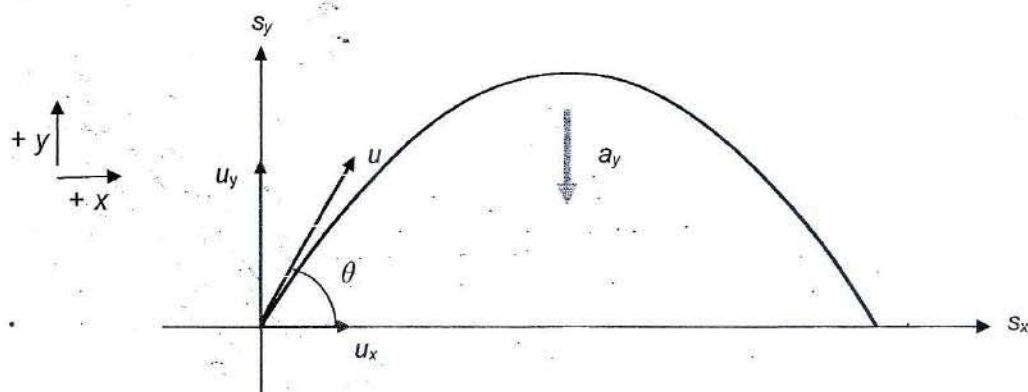
2.8.2 Projectile Motion: General Launch Angle

Let us now use our equations of motion for one-dimensional uniform acceleration motion and extend this to two-dimensional motion.

Recall our three equations of motion for one-dimensional motion:

$$\begin{aligned} v &= u + at \\ v^2 &= u^2 + 2as \\ s &= ut + \frac{1}{2}at^2 \end{aligned}$$

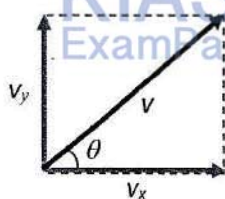
Consider an object projected above the horizontal at an angle of θ with an initial speed of u , as shown in figure below.



Analysing our projectile horizontally and vertically, and choosing our coordinate system as in the diagram, we have:

Horizontally	Vertically
$a_x = 0$	$a_y = -g$
$\Rightarrow \begin{aligned} v_x &= u_x \\ v_x &= u \cos \theta \end{aligned} \quad (1)$	$\Rightarrow \begin{aligned} v_y &= u_y + a_y t \\ v_y &= u \sin \theta - gt \end{aligned} \quad (3)$
$\Rightarrow \begin{aligned} s_x &= u_x t \\ s_x &= (u \cos \theta) t \end{aligned} \quad (2)$	$\Rightarrow \begin{aligned} s_y &= u_y t + \frac{1}{2} a_y t^2 \\ s_y &= (u \sin \theta) t - \frac{1}{2} gt^2 \end{aligned} \quad (4)$
	$\Rightarrow \begin{aligned} v_y^2 &= u_y^2 + 2 a_y s_y \\ v_y^2 &= u^2 \sin^2 \theta - 2 g s_y \end{aligned}$

In addition, the **resultant velocity** v at any instant of time can be found from the horizontal and vertical component of the velocities:



Magnitude of the velocity, $v = \sqrt{v_x^2 + v_y^2}$

Direction of the velocity θ can be determined from:

$$\tan \theta = \frac{v_y}{v_x}$$

**IMPORTANT
TO KNOW**

In exam questions, it is essential to give both the magnitude and direction for any vector quantity.

From the above equations, we can easily use them to obtain other information of the projectile, like:

- **Maximum Height:** Maximum vertical displacement with respect to the initial starting point (vertical velocity, $v_y = 0 \text{ m s}^{-1}$).
- **Horizontal Range:** Maximum horizontal displacement with respect to the initial starting point.
- **Angle of Projection for Maximum Horizontal Range.**
- **Equation of the trajectory of the projectile.**

TIPS Solving Projectile Motion Problems:

In solving problems involving projectile motion, follow the following procedures:

1. Read the question carefully.
2. Draw a careful diagram summarising the given information.
3. Choose an x-y coordinate system.
4. Identify the accelerations in the x and y directions.
5. If you are given the initial velocity, you may want to resolve it into its x and y components.
6. Analyse the horizontal and vertical motion separately.
7. Apply the relevant equations. Remember to take sign conventions into consideration.

We shall now work through several examples of projectile motion to familiarize ourselves with the strategy and the application of the equations.

Example 10

A long-jumper leaves the ground at an angle of 20° to the horizontal and at a speed of 11 m s^{-1} .

- (a) How far does he jump?
(b) What is his maximum height reached?

a) ($\rightarrow$) $a_x = 0$, $s_x = u_x t = (11 \cos 20^\circ) t$

($\uparrow$) At P,
 $s_y = 0 \text{ m}$
 $a_y = -g = -9.81 \text{ m/s}^2$
 $u_y = 11 \sin 20^\circ \text{ m/s}$

Using $s_y = u_y t + \frac{1}{2} a_y t^2$
 $\Rightarrow 0 = (11 \sin 20^\circ) t + \frac{1}{2} (-9.81) t^2$
 $t = 0 \text{ s or } t = 0.767 \text{ s}$
 (n.a)

b) At maximum height,
 ($\uparrow$) $a_y = -g = -9.81 \text{ m/s}^2$
 $u_y = 11 \sin 20^\circ \text{ m/s}$
 $v_y = 0$
 $s_y = ?$

$s_x = (11 \cos 20^\circ)(0.767 \text{ s})$
 $= 7.93 \text{ m}$

$v_y^2 = u_y^2 + 2a_y s_y$
 $s_y = \frac{0 - (11 \sin 20^\circ)^2}{2(-9.81)} = 0.72 \text{ m}$

Example 11

A stone is thrown from the top of a building upward at an angle of 30° to the horizontal with an initial speed of 20 m s^{-1} . If the height of the building is 45 m ,

- how long is the stone in flight?
- what is the speed of the stone just before it strikes the ground?

(i). Using $s_y = u_y t + \frac{1}{2} a_y t^2$

$$\Rightarrow (-45) = (20 \sin 30^\circ) t + \frac{1}{2} (-9.81) t^2$$

$$\Rightarrow 4.905 t^2 - 10 t - 45 = 0$$

Either $t = -2.915$ or $t = 4.223$
(n.a)

(ii). Using $v_y^2 = u_y^2 + 2a_y s_y$

$$\Rightarrow v_y^2 = (20 \sin 30^\circ)^2 + 2(-9.81)(-45)$$

Either $v_y = -31.4 \text{ m/s}$ or $v_y = 31.4 \text{ m/s}$
(n.a)

$$v = \sqrt{v_x^2 + v_y^2}$$

$$v = 35.9 \text{ m/s}$$

Exercise

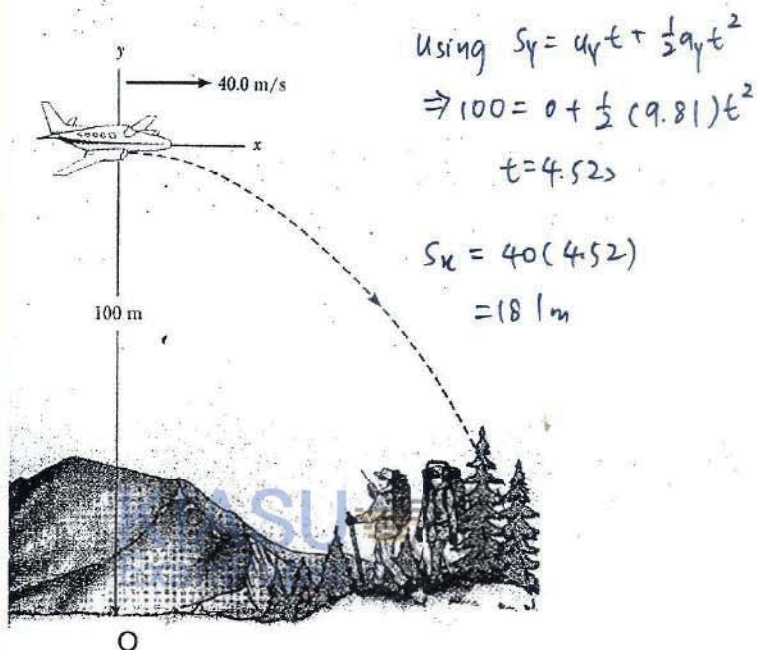
Where does the stone strike the ground?

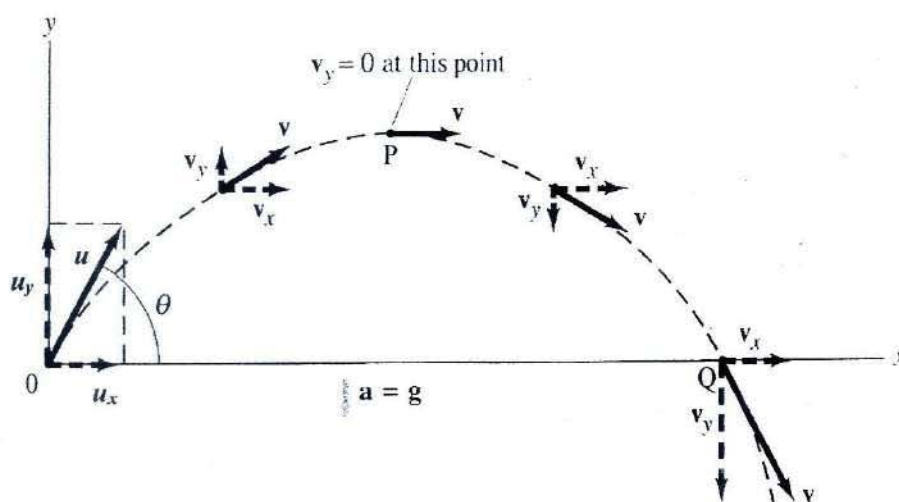
Answer

73.0 m from the base of the building.

Example 12

An Alaskan rescue plane drops a package of emergency rations to a stranded party of explorers, as shown in the figure below. If the plane is travelling horizontally at 40.0 m s^{-1} and is 100 m above a point O on the ground, how far from the point O does the package strike the ground?



Example 13: More about Projectile Motion

Consider a projectile projected with an initial velocity u at an angle above the horizontal. (Diagram above) Assuming that air resistance is negligible and the acceleration of free fall g remains constant, we can derive using the equations of motion expressions of:

- (i) maximum height H reached by the projectile,
- (ii) duration of time of flight t_{flight} ,
- (iii) horizontal range R ,
- (iv) optimal angle of projection θ_{max} (for maximum range)
- (v) equation of the trajectory of the projectile.

Adopt the sign conventions as given in the diagram above.

Solution:**(i) Maximum Height, H**

The maximum height reached is the maximum vertical displacement reached by the projectile (Point P on the diagram above). At the point of maximum height, the vertical component of the velocity is zero, hence we have

Taking upward direction as positive, consider vertical motion

$$s_y = H$$

$$a_y = -g$$

$$u_y = u \sin \theta$$

$$v_y = 0$$

Using

$$v_y^2 = u_y^2 + 2a_y s_y$$

$$\Rightarrow 0 = (u \sin \theta)^2 + 2(-g)(H)$$

$$\Rightarrow H = \frac{u^2 \sin^2 \theta}{2g}$$

(ii) Duration of Flight, t_{flight} :

The time of flight of the projectile is the time the projectile spends in the air, i.e. the time interval between the point the projectile leaves the point of projection O to the point where it lands on the ground Q.

At point Q, the vertical displacement of the projectile from point O is zero. Hence we have Taking upward direction as positive, consider vertical motion,

$$s_y = 0$$

$$a_y = -g$$

$$u_y = u \sin \theta$$

Using $s_y = u_y t + \frac{1}{2} a_y t^2$

$$\Rightarrow 0 = (u \sin \theta) t_{\text{flight}} + \frac{1}{2} (-g) t_{\text{flight}}^2$$

$$\Rightarrow t_{\text{flight}} = \frac{2u \sin \theta}{g} \quad \text{or} \quad t_{\text{flight}} = 0 \quad (\text{at the time of beginning of throw when } s_y = 0)$$

(iii) Horizontal Range, R :

Since there are no horizontal forces acting on the projectile, the horizontal velocity remains constant during the time of flight and the horizontal acceleration is zero. Using, the result for the time of flight that we derived in (ii), hence, we have

Consider horizontal motion,

$$v_x = u_x = (u \cos \theta)$$

$$a_x = 0$$

$$t_{\text{flight}} = \frac{2u \sin \theta}{g}$$

$$s_x = R$$

Using $s_x = u_x t + \frac{1}{2} a_x t^2$

$$\Rightarrow R = (u \cos \theta) \left(\frac{2u \sin \theta}{g} \right) + 0$$

But $2 \cos \theta \sin \theta = \sin 2\theta$

Hence, $R = \frac{u^2 \sin 2\theta}{g}$

(iv) Optimal Angle of Projection θ_{max} (for maximum range):

Now, $R = \frac{u^2 \sin 2\theta}{g}$

But $-1 \leq \sin 2\theta \leq 1 \Rightarrow \text{maximum } \sin 2\theta = 1$

Therefore, the maximum possible range, $R_{\text{max}} = \frac{u^2}{g}$

This occurs when $\sin 2\theta = 1 \Rightarrow 2\theta = 90^\circ \Rightarrow \theta = 45^\circ$

(v) The Trajectory Equation:

To show that the path followed by any projectile is a parabola, we need to find the y as a function of x by eliminating t between the vertical motion and the horizontal motion.

We can write following equations for both the horizontal and vertical displacements:

$$x = u_x t \quad (1)$$

$$y = u_y t - \frac{1}{2} g t^2 \quad (2)$$

From (1), we have $t = x/u_x$, and we substitute it into (2) to obtain,

$$y = \left(\frac{u_y}{u_x}\right)x - \left(\frac{g}{2u_x^2}\right)x^2$$

If we write $u_x = u \cos \theta$ and $u_y = u \sin \theta$, we can therefore write

$$y = (\tan \theta)x - \left(\frac{g}{2u \cos^2 \theta}\right)x^2$$

We can see that y as a function of x has the form

$$y = ax - bx^2$$

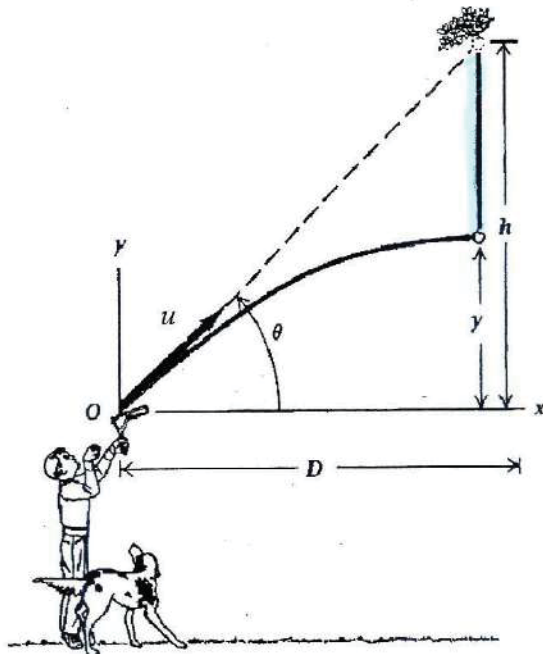
This is an equation of a parabola.



We do NOT recommend you to memorise the formulae derived in this example. During examinations, you will NOT be given any credit for simply quoting these results directly without showing derivation.

Challenging Problem: Example 14

A boy aims his slingshot directly at an apple hanging in a tree. At the moment he shoots a pebble, the apple drops. If pebble hits target, both pebble and target (i.e. the apple) are at SAME location at the SAME time. Horizontal distance, D , travelled by stone is known. Thus time t , time for the pebble to travel the distance D can be determined. Show that at time t , apple is at the same height as pebble.

**Solution**

For apple, vertical displacement from slingshot,

$$(\downarrow)s_{apple} = h - (u_y t + \frac{1}{2} a_y t^2)$$

$$s_{apple} = h - (0 + \frac{1}{2} g t^2)$$

$$s_{apple} = h - \frac{1}{2} g t^2 \text{ -----(1)}$$

For pebble,

$$\text{horizontal displacement from slingshot, } (\rightarrow)s_{pebble(x)} = u \cos \theta \times t = D \text{ -----(2)}$$

$$\text{vertical displacement from the slingshot, } (\uparrow)s_{pebble(y)} = u_y t + \frac{1}{2} a_y t^2$$

$$(\uparrow)s_{pebble(y)} = u \sin \theta \times t - \frac{1}{2} g t^2 \text{ -----(3)}$$

(3) / (2):

$$s_{pebble(y)} = (\tan \theta - \frac{1}{2} g t^2 / u \cos \theta \times t) D$$

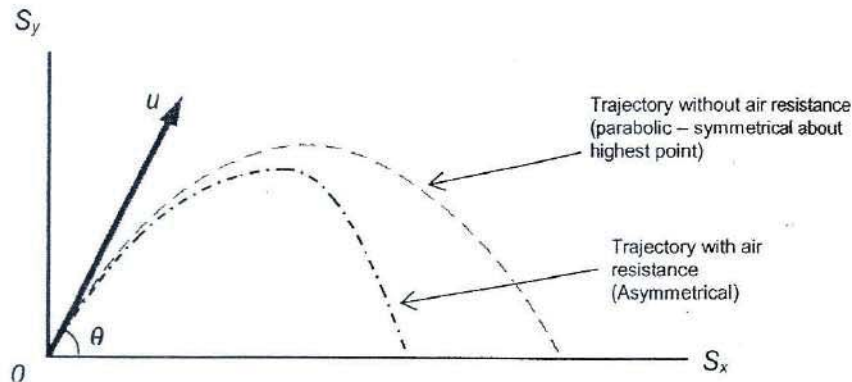
$$\text{Since } h = \tan \theta \times D$$

$$s_{pebble(y)} = h - \frac{1}{2} g t^2 \text{ -----(4)}$$

Equations (1) and (4) are identical, thus both apple and pebble is at the same height at the same time so the pebble will hit the apple.

2.8.3 Projectile Motion with Air Resistance

Consider an object projected above the horizontal with an initial speed of u and at angle of θ with respect to the horizontal.



Suppose that the air resistance experienced by the projectile is not negligible. Then **throughout the motion, the projectile will experience a drag force in opposite direction to its motion** (velocity).

Let us further analyse the horizontal motion and vertical motion independently,

Horizontally,



The drag force acts in opposite direction to the horizontal velocity.
 $\Rightarrow$ The horizontal velocity reduces with time.
 $\Rightarrow$ The **horizontal range is reduced** (as compared with the case if no air resistance present).

Vertically,

On its flight up,



The drag force acts in the same direction as the weight mg .

The magnitude of the net acceleration is greater than g .

$\Rightarrow$ **Maximum height reached** is lower than if there were no air resistance present.

KIASU
ExamPaper



On its flight down,



The drag force acts in opposite direction to the weight mg .

The magnitude of the net acceleration is less than g .

$\Rightarrow$ The final (vertical) speed v_y will be smaller than the initial (vertical) speed u_y (at the point of projection).

$\Rightarrow$ The average (vertical) speed upwards is greater than the average (vertical) speed downwards.

Since, distance travelled up = distance to come down
and Average speed = Total distance / Time taken

$\Rightarrow$ **The time taken to travel up is shorter than the time taken to travel down.**

Furthermore, comparing the vertical motion of the projectile on its flight up and on its flight down, we see that:

- The (vertical) acceleration experienced by the projectile when it is going up is different from the (vertical) acceleration experienced by the projectile when it is coming down. ($a_{y,up} \neq a_{y,down}$)
- For a particular height, the speed of the projectile when it is going up is also different for the speed of the projectile when it is coming down.

Hence, we can see that the **path of the projectile will not be symmetrical about the highest point**, i.e. the path is asymmetrical about the highest point.

In summary,

The key characteristics of a projectile motion with air resistance (compared to if air resistance is negligible) are:

1. Horizontal range is shorter.
2. Maximum height is lower.
3. The time taken to reach maximum height must be shorter than the time taken to return to the starting point.
4. Path is asymmetrical about a vertical line through its highest point.

Further Readings and References

Appendix I : Understanding the Calculus Notations

Although rigorous treatment with calculus is not required in the A-level syllabus, it is however useful for us to understand the physical significance of some of the calculus notations as they are commonly used to define physical quantities.

The Derivative : $\frac{dy}{dx}$

Consider a graph of y vs. x as indicated by the curve in Figure 1.

1. $\frac{\Delta y}{\Delta x}$ indicates the gradient of the line joining the 2 points A and B on the line.

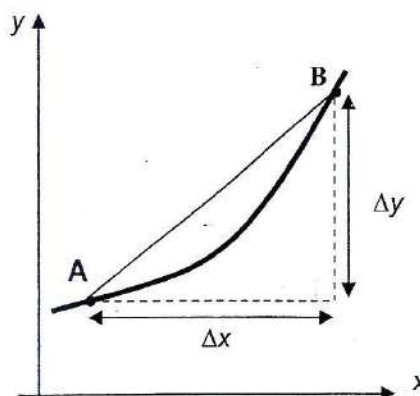


Figure 1. Graph of y vs. x

Figure 2 shows that if we move point B closer towards point A, eventually the 2 points will meet. At this point gradient of the line AB will be exactly equal to the gradient of the tangent of the curve at A. (as represented by the bold straight line). Mathematically, we say:

$$\lim_{\Delta x \rightarrow 0} \frac{\Delta y}{\Delta x} = \frac{dy}{dx} = \text{gradient of the tangent of the } y - x \text{ graph}$$

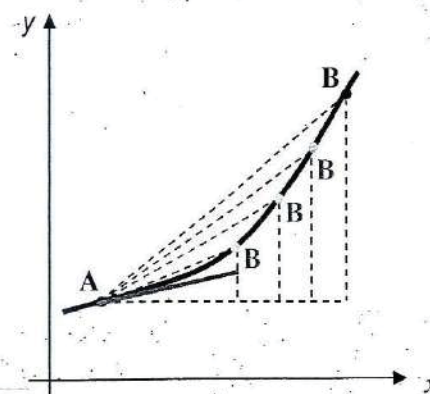


Figure 2. Tangent of the curve at A.

This physical interpretation is very important in physics as many physical quantities are expressed in this derivative form.

For example, in our study of kinematics, we have :

Velocity, v = Rate of change of displacement with respect to time

$$= \frac{ds}{dt}$$

= Gradient of tangent of the $s - t$ graph

and

Acceleration, a = Rate of change of velocity with respect to time

$$= \frac{dv}{dt}$$

= Gradient of the tangent of the $v - t$ graph

The Integral: $\int y \, dx$

The reverse process to taking the derivative is referred to either as *integration* or finding the *anti-derivative*. Graphically, it is equivalent to finding the area under the curve. Let us see why.

Consider a $y - x$ graph as shown in Figure 3. Suppose we wish to find the area under the graph from some initial x value, x_{initial} , to some final x value, x_{final} . One way to do so is break the area up into n number of small little rectangles each of some area, ΔA .

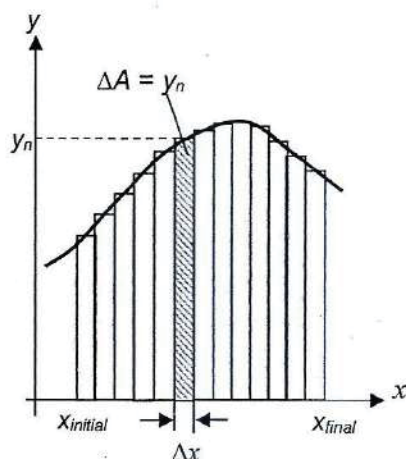


Figure 3. Finding the area under a curve. The area of the shaded rectangle is equal to the product of the width of the rectangle Δx and the y value of the curve at the particular x , y_n .

Hence, the area of each small rectangle can be written as : $\Delta A = y_n \Delta x$

and the total area under the curve can be estimated by summing the area of all these little rectangles, i.e.

$$\text{Total area } A \approx \sum_n y_n \Delta x$$

where the symbol Σ (upper case Greek sigma) signifies a sum over all terms, i.e. over all values of n .

For the approximation of the area to be even better, we can try to divide the area into smaller and smaller rectangles, by decrease the interval Δx , i.e. we say that we let $\Delta x \rightarrow 0$.

Therefore,

$$A = \lim_{\Delta x \rightarrow 0} \sum_n y_n \Delta x$$

This is also written as $\int_{x_{\text{initial}}}^{x_{\text{final}}} y \, dx$, therefore $\int_{x_{\text{initial}}}^{x_{\text{final}}} y \, dx = \text{Area under the } y - x \text{ graph}$

We now use the defining equations for acceleration and velocity to illustrate the use of the above concept.

The defining equation for acceleration of an object is $a = \frac{dv}{dt}$

Therefore, the change in velocity of an object is $v_{\text{final}} - v_{\text{initial}} = \int_{t_{\text{initial}}}^{t_{\text{final}}} a \, dt = \text{Area under } a - t \text{ graph}$

Similarly, velocity of an object is $v = \frac{ds}{dt}$

Rearranging the equation, we also have the change in displacement of an object given as

$$s_{\text{final}} - s_{\text{initial}} = \int_{t_{\text{initial}}}^{t_{\text{final}}} v \, dt = \text{Area under } v - t \text{ graph}$$

Contents

H2 Physics Syllabus 9749	3
3.0 Introduction	3
3.0.1 Types of Forces	4
3.0.2 Weight	5
3.0.2.1 Centre of Gravity	5
3.0.3 Contact Force	6
3.0.3.1 Normal Contact Force	6
3.0.4 Friction	6
3.0.5 Tension	7
3.1 Newton's Laws of Motion	8
3.1.1 Newton's 1st Law	8
3.1.2 Linear Momentum	8
3.1.3 Newton's 2nd Law	10
3.1.4 Impulse	12
3.1.5 Newton's 2 nd law for constant mass	14
Net force vector	16
Applying $F_{net} = ma$ to a particle (revision) with forces in 1D	17
Applying $F_{net} = ma$ to a particle with forces in 2D	20
3.1.6 Newton's 3 rd Law	21
3.2 Principle of Conservation of Linear Momentum	25
3.2.1 Two-body System	26
3.2.2 Types of Collisions	27
3.2.2.1 Head-on vs Glancing Collisions	27
3.2.2.2 Elastic vs Inelastic Collisions	27
3.2.2.3 Relative Speed of Approach/Separation during Elastic Collision	29
Appendix A: Static and Kinetic Friction	34
Appendix B: Centre of Mass	35
Appendix C: Coefficient of Restitution	36
Appendix D: Newton's Law in Car Safety	36

H2 Physics Syllabus 9749

Topic 3: Dynamics

Dynamics	Learning Outcomes Students should be able to:
Newton's laws of motion	(a) state and apply each of Newton's laws of motion. (b) show an understanding that mass is the property of a body which resists change in motion (inertia). (c) describe and use the concept of weight as the effect of a gravitational field on a mass.
Linear momentum and its conservation	(d) define and use linear momentum as the product of mass and velocity. (e) define and use impulse as the product of force and time of impact. (f) relate resultant force to the rate of change of momentum. (g) recall and solve problems using the relationship $F=ma$, appreciating that resultant force and acceleration are always in the same direction. (h) state the principle of conservation of momentum. (i) apply the principle of conservation of momentum to solve simple problems including inelastic and (perfectly) elastic interactions between two bodies in one dimension. (Knowledge of the concept of coefficient of restitution is not required.) (j) show an understanding that, for a (perfectly) elastic collision between two bodies, the relative speed of approach is equal to the relative speed of separation. (k) show an understanding that, whilst the momentum of a closed system is always conserved in interactions between bodies, some change in kinetic energy usually takes place.

Topic 4: Forces (Part)

Forces	Learning Outcomes Students should be able to:
Types of force	(b) describe the forces on a mass, charge and current-carrying conductor in gravitational, electric and magnetic fields, as appropriate. (c) show a qualitative understanding of normal contact forces, frictional forces and viscous forces including air resistance. (No treatment of the coefficients of friction and viscosity is required.)
Centre of gravity	(d) show an understanding that the weight of a body may be taken as acting at a single point known as its centre of gravity.

3.0 Introduction

Mechanics is the branch of Physics concerned with the motion of objects subjected to forces.

3.0.1 Types of Forces

Broadly speaking, forces may be categorized into two kinds: **contact** forces and **non-contact** forces.

1. Contact forces

Forces acting between two bodies that are in physical contact.

Contact forces	Description
Friction	Force that resists the relative motion or the tendency for relative motion between two surfaces that are in contact.
Normal Contact Force	Net electrostatic repulsion between two surfaces that are in contact.
Contact Force	Vector sum of friction and normal contact force, meaning the total contact force.
Spring Force	Force exerted by one end of a spring when the spring is extended or compressed from its natural length.
Tension	Force exerted by one end of a string/rope which is being pulled taut.
Upthrust	Net upward force exerted by the fluid on a body fully or partially submerged in the fluid.
Viscous drag force	Force that resists a body moving relative to a fluid.
Air resistance	Drag force exerted by air on a body moving in air.

2. Non-contact forces

Forces acting between two bodies which are NOT in physical contact (separated by a distance). This characteristic of the force is described as **action-at-a-distance**.

Non-contact forces	Description	Physical property of a body on which the force acts on.
Gravitational force	Attractive force exerted between masses.	Mass
Electric force	Attractive/repulsive force exerted between electrically charged bodies.	Electric charge
Magnetic force	Force exerted between currents	Moving charges

3.0.2 Weight

The weight of a body is due to the body being in the gravitational field of a massive body. The gravitational force exerted on the body by Earth is called its weight (due to Earth).

Notes:

- Weight is a **force** (hence a **vector**).
- Direction: Weight acts on the centre of gravity of the object towards the centre of mass of the body producing the gravitational field. For everyday-life problems, the weight of the object is taken to act **vertically downwards**.
- Magnitude: $W = mg$
 W : weight of an object (SI unit: N)
 m : mass of the body (SI unit: kg)
 g : acceleration of free fall/gravitational field strength at the point (SI unit: m s^{-2})

Distinction between mass and weight

Mass	Weight
A measure of the inertia of a body	A measure of the gravitational pull on its mass
Scalar quantity	Vector quantity
SI unit: kg	SI unit: N
Same regardless of location	Varies with gravitational field strength at each location
Same regardless of the method of measurement.	Some methods of measurement do not measure the true weight but the apparent weight.



In everyday language, the terms 'mass' and 'weight' are used interchangeably, e.g. I "weigh" 70 kg. In physics, and especially during examinations, such casual usage is not tolerated.

3.0.2.1 Centre of Gravity

An object of a significant size may have unequal distribution of mass.

The **centre of gravity** of an object is the point at which the weight of the object *appears to act*.

Notes:

- The force due to weight is drawn from the centre of gravity of an object.
- The centre of gravity of an object is not necessarily on the object. It may lie outside the object.
- In a uniform gravitational field, the centre of gravity always coincides with the centre of mass of the object. For a mathematical definition on centre of mass, refer to Appendix B.

3.0.3 Contact Force

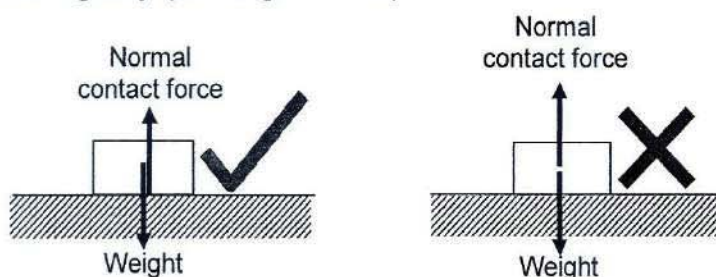
When two surfaces come into contact, they both experience contact forces arising from electrostatic forces between atoms of the surfaces. For ease of analysis, this contact force is often resolved into two components: The component perpendicular to the surfaces is called **normal contact force**, and the component parallel to the surfaces is given the name **friction**.

3.0.3.1 Normal Contact Force

A block is able to rest on a table because its weight is balanced by an upward force the table exerts on it. This force arises from the electrostatic repulsion between the molecules of the table and the molecules of the block. It is called the normal contact force since it acts perpendicularly (normal) to the surfaces.



In free body diagrams (see Section 3.5.1 for detail), contact forces must be drawn from a contact point (e.g. somewhere along the base of the block). A common mistake is a student draws them from the centre of gravity. (See diagram below)



- By definition, the normal contact force is always directed perpendicular to the surfaces.
- The magnitude depends on how hard the two surfaces are pressed into each other.

3.0.4 Friction

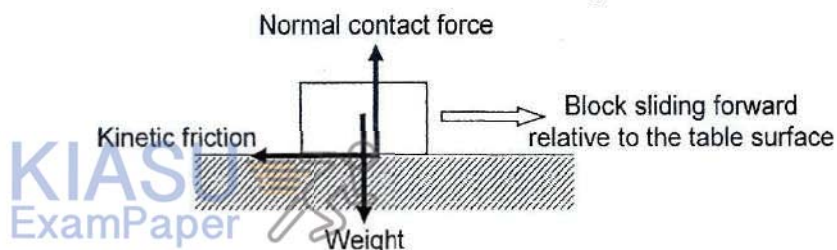
Definition of Friction

Friction is the force that resists relative motion between two sliding surfaces (called kinetic friction), or the tendency for relative motion between two stationary surfaces (called static friction).

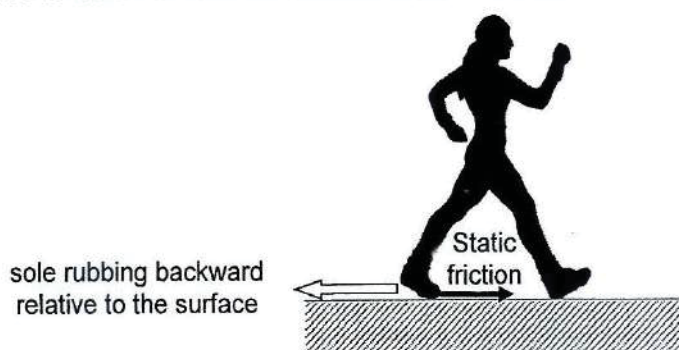
Direction

Friction is always directed in opposite direction to the direction of relative motion.

For example, a block sliding along a rough table is gradually decelerated to rest. It experiences a (kinetic) friction directed backward because the block's surface is sliding forward relative to the table's surface.



On the other hand, a person is able to walk forward only if he experiences a (static) friction that is directed forward. The sole of his shoe tends to rub backward relative to the floor's surface.



Magnitude of Friction (Good to know, Not stated in syllabus)

The magnitude of friction between two surfaces follows these empirical rules:

1. It is dependent on the *nature of the surface*. For example, friction between glass and rubber is usually larger than friction between glass and wood.
2. It is proportional to the *normal contact force* between the two surfaces, which in turn depends on how hard the two surfaces are pressed against each other.
3. It is independent of the area of contact between the two bodies.

Mathematically, the magnitude of friction can be expressed as

$$f \leq \mu N$$

f : friction

μ : coefficient of friction between the two surfaces.

N : normal contact force

It is interesting that the coefficient of friction between two stationary surfaces (tending to have relative motion) is always larger than the coefficient of friction of the same surfaces when they are in relative motion. For more detail, refer Appendix A.

3.0.5 Tension

In physics, tension refers to the magnitude of the pulling force exerted by a string, rope, cable or chain on the object tied to its end when it is pulled **taut**.

- The direction of the force exerted on the object is along the string, directed away from the object.
- The magnitude of the force depends on how much the string is stretched. Once the string is **slack**, the tension falls to zero.
- In many situations, we idealise the string, rope, cable, etc to be **massless** and **inextensible** in order to simplify analysis.
- Strings, ropes, etc are often used together with pulleys. Again, in many situations, we idealise a pulley to be **massless** and **frictionless** (at the axle) in order to simplify analysis.

Consider an object which is at rest, is hung vertically by a rope.



3.1 Newton's Laws of Motion

3.1.1 Newton's 1st Law

Statement of Newton's 1st law of motion

A body stays at rest or continues to move with a constant speed in a straight line unless acted upon by a **net external force**.

Notes

- Changes in a body's state of motion refer to changes in its velocity. It can be a change in either its **speed** or **direction** (or both).
- An object's resistance to change its state of motion is called **inertia**. As such Newton's 1st law is also referred to as the Law of Inertia.
- An object's inertia is dependent on its mass: The larger its mass, the larger its inertia.

3.1.2 Linear Momentum

Everyday experience tells us that 2 objects moving at the same velocity can result in very different consequences. Consider a 3-gram ping pong ball and a 3-ton lorry, both travelling at 5 m s^{-1} . Which would you rather collide with?

In Physics, we define a "quantity of motion" called **momentum, p** .

Definition of Linear momentum

The linear momentum of a body is the product of its mass and its velocity.

Mathematically, linear momentum, $\vec{p} = m\vec{v}$

Notes

- Momentum is **vector quantity**. It shares the same direction as the object's velocity.
- Momentum has the SI unit: **kg m s^{-1}**

Example 1a

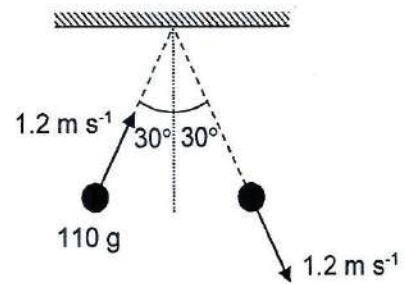
A 110 g billiard ball rebounds off the wall with velocities as shown in the diagram. The ball moves on a horizontal plane. Evaluate the change in momentum of the ball.

$$\vec{\Delta p} = \vec{p}_f - \vec{p}_i = m(\vec{v}_f - \vec{v}_i)$$

$$\Delta v = \sqrt{1.2^2 + 1.2^2 - 2(1.2)(1.2)\cos(20^\circ)}$$

or

$$\vec{\Delta p} = 0.100 \times \Delta v$$



3.1.3 Newton's 2nd Law

Statement of Newton's 2nd law of motion

The **rate of change** of linear momentum of a body is **directly proportional** to the **resultant force** acting on it and its direction is in the **same direction** as this resultant force.

Mathematically, Net force, $\vec{F}_{net} = \frac{d\vec{p}}{dt} = \frac{d(m\vec{v})}{dt}$

Average net force, $\langle F_{net} \rangle = \frac{\Delta\vec{p}}{\Delta t} = \frac{\Delta(m\vec{v})}{\Delta t}$

SI unit for force is the newton (N).

Definition of the newton

One newton is defined as the amount of force that gives an acceleration of 1 m s^{-2} to a body with a mass of 1 kilogram. (i.e. $1 \text{ N} \equiv 1 \text{ kg m s}^{-2}$)

Strictly speaking, N2L only states that $\vec{F}_{net} \propto \frac{d\vec{p}}{dt}$

$\Rightarrow F_{net} = k \frac{dp}{dt}$ where k is some constant of proportionality

However, the SI unit newton has been defined such that the constant $k = 1$ (dimensionless).

Example 1b (continued)

The time interval of contact between the billiard ball and the wall is 0.100 s. Find the average net force acting on the ball during the collision.

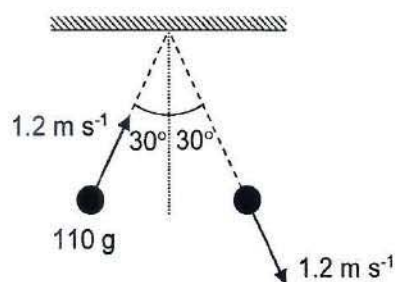
Hence, deduce the average net force acting on the wall during the collision.

By 2nd law, $\Delta p = 0.23 \text{ kg m/s}$

$\frac{\Delta p}{\Delta t} = \frac{0.23}{0.100} = 2.3 \text{ N directed } \downarrow$

By 3rd law,

$|F_{\text{on wall by ball}}| = |F_{\text{on ball by wall}}| \approx 2.3 \text{ N directed upwards}$



KIASU
ExamPaper



Note that Newton's 3rd Law is utilized in solving the above problem. By calculating the magnitude of the force that A acts on B, we indirectly obtained the magnitude of the force that B acts on A.

Example 2

The motor of a small boat has a propeller which sends back a column of water of cross sectional area 0.030 m^2 at a speed of 8.0 m s^{-1} . Assuming that the boat is held at rest, calculate

- the rate (in kg s^{-1}) at which the water is propelled backwards
- the rate of change of momentum of the water
- the thrust exerted by the motor on the boat

Density of water = $1.0 \times 10^3 \text{ kg m}^{-3}$

In 1 sec,
 $\text{Mass} = \rho AL = 240 \text{ kg}$

Mass rate, $\frac{dm}{dt} = 240 \text{ kg/s}$

$$\begin{aligned} b) \quad \frac{dp}{dt} &= \frac{\Delta p}{\Delta t} = \frac{m(v-u)}{\Delta t} = \frac{m}{\Delta t} (v-0) \\ &= \frac{240(8.0-0)}{1} = 1920 \text{ kg m/s}^2 = 1920 \text{ N} \end{aligned}$$

Rate of change of momentum of the water is actually the force exerted on the water by the motor.

- c) As the motor propels the water, it exerts a force on the water in the backward direction.

By Newton's 3rd law of motion, water exerts an equal and opp. force on the motor, in the forward direction

$$F_{\text{water on motor}} = -F_{\text{motor on water}}$$

$$\therefore \text{thrust on boat in forward direction} = 1920 \text{ N}$$

KIASU
ExamPaper



Note that Newton's 3rd Law is utilized in solving the above problem. By calculating the magnitude of the force that A acts on B, we indirectly obtained the magnitude of the force that B acts on A.

3.1.4 Impulse

Definition of Impulse

The impulse of a force is the product of the average force and the time interval over which it is applied.

Mathematically, $\vec{J} = \vec{F}\Delta t$

Notes

- Impulse is a vector quantity; its direction is the same as the force F .
- The SI units for impulse are **N s**. They can be shown to be equivalent to **kg m s⁻¹**, the same as the units of momentum

3.1.4.1 Relation between Impulse and Momentum

From Newton's 2nd Law, $\vec{F} = \frac{d\vec{p}}{dt}$

Rearranging and integrating, we obtain

$$\int \vec{F} dt = \int d\vec{p}$$

The LHS, $\int \vec{F} dt$, is the **impulse** of the force.

The RHS, $\int d\vec{p} = \vec{p}_f - \vec{p}_i = \Delta\vec{p}$, is the **total change in momentum**.

This result which links impulse to the change in momentum is known as the **Impulse-Momentum Theorem**.

Statement of Impulse-Momentum Theorem

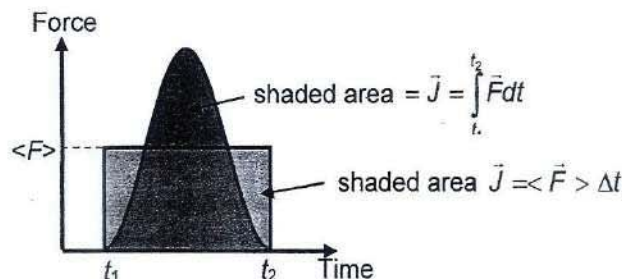
The **change in** momentum of a particle during a time interval equals the impulse of the net force that acts on the particle during that interval.

Mathematically, $\vec{J} = \int \vec{F} dt = \Delta\vec{p}$

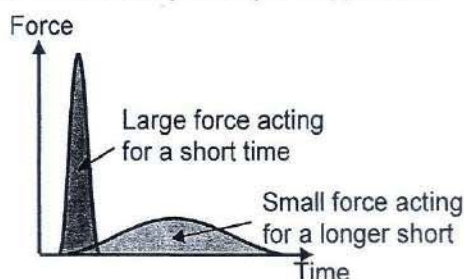
For **average** or **constant** forces, this simplifies to $\vec{J} = \langle \vec{F} \rangle \Delta t = \Delta\vec{p}$

3.1.4.2 Graphical Representation of Impulse

Since $\vec{J} = \int \vec{F} dt = \Delta \vec{p}$, impulse is represented by the **area under the force-time graph**. The same area also represents the **change in the momentum** of the body.



To illustrate further, study the $F-t$ graphs below. Notice even though the two forces have different peak magnitude, both deliver the same impulses (i.e. same area under each graph). Hence the two forces will cause the same change in momentum to the object they are applied on.

**Example 3 (N89/II/8 modified)**

In order to stop a car of mass 1500 kg travelling at 30 m s⁻¹, the driver applies his brakes so that F , the total stopping force, increases steadily to a maximum and then decreases to zero as shown in the figure. Calculate

(a) the momentum of the car when it is travelling at 30 m s⁻¹

(b) the impulse due to the braking force

(c) the magnitude of the average stopping force, $\langle F \rangle$

(d) the value of $F_{\max}$

a) $1500 \times 30 = 45 \times 10^3 \text{ kg m/s}$ in the forward direction

b) $p_f - p_i = (1500)(0 - 30) = -45 \times 10^3 \text{ kg m/s}$
opposing the car's motion

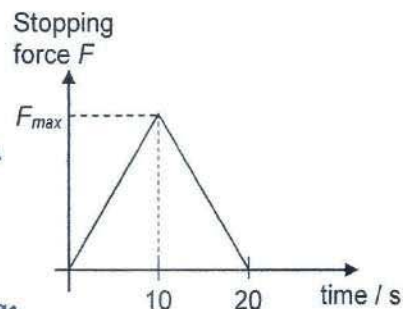
c) $\langle F \rangle \Delta t = \Delta p_{\text{car}}$

$\langle F \rangle = \frac{\Delta p_{\text{car}}}{\Delta t} = \frac{-45 \times 10^3}{20} = -2.3 \times 10^3 \text{ N}$

$|\langle F \rangle| = 2.3 \times 10^3 \text{ N}$

d) $\Delta p_{\text{car}} = \text{Area under } F-t \text{ graph}$

$45 \times 10^3 = \frac{1}{2} F_{\max} (20)$



$F_{\max} = 4500 \text{ N}$

3.1.5 Newton's 2nd law for constant mass

From N2L,
$$\vec{F}_{net} = \frac{dp}{dt} = \frac{d(m\vec{v})}{dt}$$

For constant mass,
$$\vec{F}_{net} = m \frac{d\vec{v}}{dt}$$

$$\vec{F}_{net} = m\vec{a}$$

Notes

- $\vec{F}_{net} = m\vec{a}$ is **not** a complete representation of Newton's 2nd Law. It only represents a special case of Newton's 2nd law for constant mass.
- The product $m\vec{a}$ represents the **resultant** of all the actual physical forces acting on a body. It is **not** a new or additional physical force. As such, we do not draw $m\vec{a}$ as a separate force in free body diagrams.
- Be clear about which is the cause ($\vec{F}_{net}$), and which is the effect (acceleration).
 - A non-zero $\vec{F}_{net}$ *results in* acceleration (change in speed or direction), or
 - If acceleration is detected, it *implies* $\vec{F}_{net}$ is non-zero. It does not result in a new force.

TIPS Solving $\vec{F}_{net} = m\vec{a}$ Problems

Step 0 Illustrate if no diagram is provided.

Step 1 Consider the expected direction of acceleration and resultant force.
 (This is to facilitate setting up the $\vec{F}_{net} = m\vec{a}$ equation.)

Step 2 Draw the FBD for the body(ies) of interest.

Step 3 Establish the coordinate system, with one axis parallel to the expected acceleration, and a mutually perpendicular axis if necessary.

Step 4 Resolve the forces into components parallel to your chosen axis/axes.

Step 5 Apply $\vec{F}_{net} = m\vec{a}$ and solve for all unknowns.

(You need as many independent equations as there are unknowns to solve them uniquely)

3.1.5.1 Free Body Diagram

A free body diagram (FBD) is a simple representation of an object and a vector representation of all the forces (including all contact and non-contact forces) acting on the object.

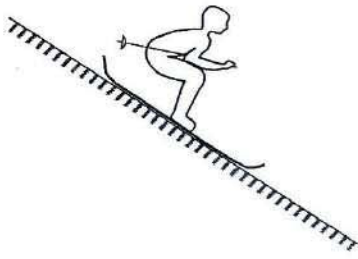
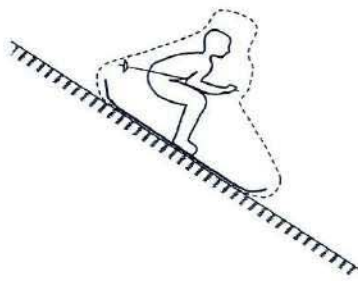
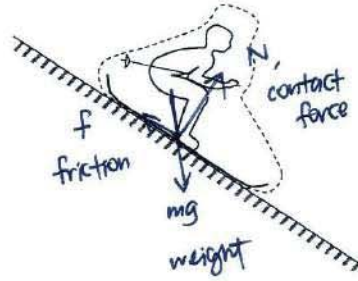
TIPS Drawing Free Body Diagram

1. Draw a simplified diagram if one is not provided.
2. Identify and isolate the body / bodies (or system) of interest.
3. Draw **all** the forces acting **on** the body/ bodies (or system) using labelled arrows.
4. The forces must be drawn from the point of application of the forces.
5. Draw forces to appropriate lengths since the relative lengths of the arrows represent the relative magnitude of the forces.
6. For two-dimensional problems, resolving forces into two mutually perpendicular directions can significantly simplify analysis.
 - E.g. horizontal and vertical direction
 - E.g. parallel and perpendicular to a slope.

Note: the resolved components must be drawn in **dotted** lines so as to distinguish them from the actual physical force.
7. Label the arrows clearly with words or defined symbols.

In this section, you will learn how to construct a free-body diagram correctly.

In Exercise 1, draw the free-body diagram for a skier going downhill at constant speed. Assume air resistance is negligible. (This is done for you except 1c). Indicate the external forces acting on the body or system as RED arrows at the suitable places.

1a. A snapshot or photo of the system of interest. Draw the diagram if it is not provided.	
b. Identify and isolate the body / bodies of interest by drawing a boundary around it	
c. Draw the external forces acting on the body, as labelled arrows starting from the point of application of force. - the contact forces wherever the boundary touches another external body - the non-contact force (weight, electric or magnetic force) Ensure the lengths of the arrows are proportional to the magnitude of the force	

Net force vector

In this section, we will construct the net force vector acting on a body by recognising the change in velocity. The vector sum of external forces acting on the body is the net force acting on the body.

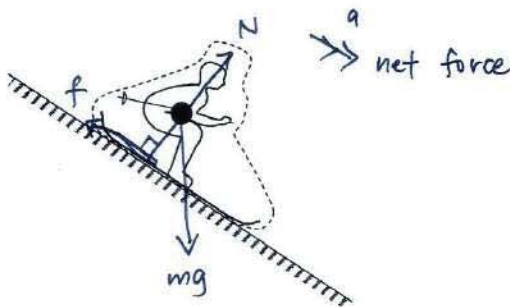
When the velocity is changing, a net force is required. What is the direction of the net force vector?

The net force is in the direction of the acceleration / change in velocity.

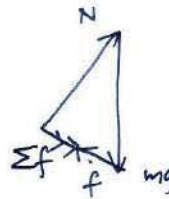
In exercises 2 to 3,

- draw the net force vector as a GREEN arrow with correct length.
 - draw a vector diagram to show the relationship between the external forces (in red) and the net force (in green)
- You will need to approximate the skier as a single particle (black dot) at the centre of mass.

2. a) Draw the net force vector for the skier speeding up

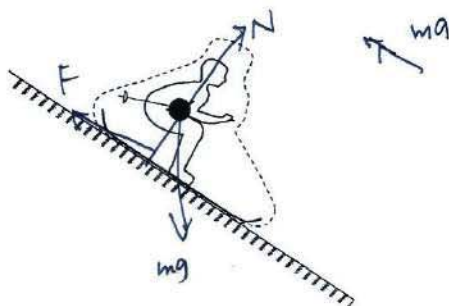


b) vector diagram with the net force

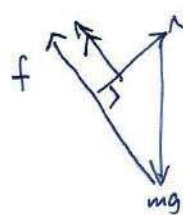


$$\begin{aligned} & \text{By } N \perp L, \\ & \Sigma f = ma \\ & \vec{N} + \vec{f} \end{aligned}$$

3. a) Draw the net force vector for the skier slowing down



b) vector diagram with the net force



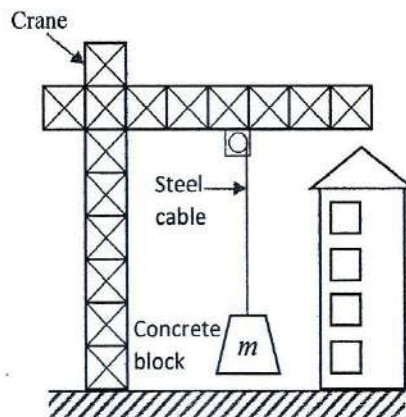
$$\begin{aligned} & \text{By } N \perp L, \\ & \Sigma f = ma \\ & \vec{N} + \vec{mg} + \vec{f} = \vec{ma} \end{aligned}$$

Applying $F_{net} = ma$ to a particle (revision) with forces in 1D

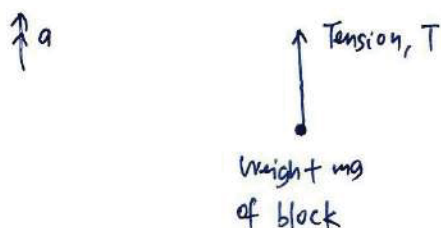
In Examples 4 to 6 we will apply the simplified form of Newton's 2nd law ($F_{net} = ma$) to write down the equation for the forces acting on a particle in 1D.

Example 4

A crane lifts a concrete block ($m = 500 \text{ kg}$). The concrete block increases in speed at a uniform rate of 5.0 m s^{-2} after it leaves the ground. Assume air resistance is negligible. Find the tension in the steel cable.



- a) Draw the free-body diagram for the concrete block (m) after it leaves the ground.



- b) Write down the equation for the forces acting on m , and determine the tension in the steel cable.

By N2L,

$$(\uparrow+) \quad F_{net} = ma$$

$$T - mg = ma$$

$$T = mg + ma$$

$$= m(g + a)$$

$$= 500(9.81 + 5.00)$$

$$= 7400$$

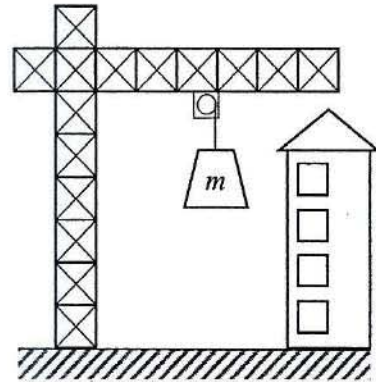
KIASU
ExamPaper



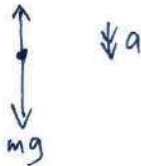
[7400 N]

Example 5

5. A crane lifts a concrete block. The concrete block slows down at a uniform rate of 5.0 m s^{-2} as it approaches the top of the building. Assume air resistance is negligible. Find the tension in the steel cable.



a) Draw the free-body diagram for the concrete block (m) as it approaches the top of the building.



b) Write down the equation for the forces acting on m and determine the tension in the steel cable.

By N2L,

$$(\downarrow+) \quad F_{\text{net}} = ma$$

$$mg + (-T) = ma$$

$$T = mg - ma$$

$$= 500(9.81 - 5.00)$$

$$= 2400 \text{ N}$$

Now, try to do Example 6 on your own. Assume air resistance is negligible.

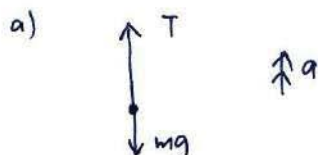
Example 6

A 1000-kg helicopter accelerates upwards at 2.5 m s^{-2} . Determine the upwards thrust, T , generated by the helicopter blades.

a) Draw the free-body diagram for *the helicopter*.

b) Write down the equation for the forces acting on *the helicopter* to determine the upwards thrust.

[12000 N]

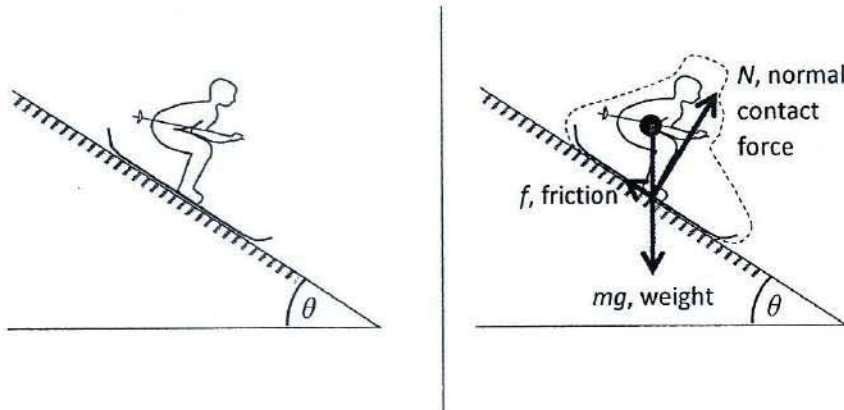


b) $(\uparrow +) \quad \Sigma F = ma$
 $T + (-mg) = ma$
 $T = m(a + g)$
 $= 1000(2.5 + 9.81)$
 $= 12000 \text{ N}$

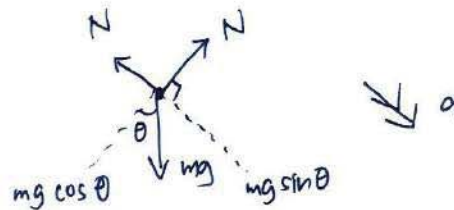
Applying $F_{\text{net}} = ma$ to a particle with forces in 2D**Example 7**

We revisit the downhill skier and apply $F_{\text{net}} = ma$ on a particle in two perpendicular directions. The 75-kg skier is accelerating at 2.6 m s^{-2} down a slope at an angle $\theta = 30^\circ$. Find the friction f and the normal contact force N acting on the skier.

a) Draw the free-body diagram of the skier (already drawn for you).



b) Translate the force vectors such that they all act from a single point which is the centre of mass of the skier (at the black dot).



b) Write down the equations for the forces acting on the skier.

$$\rightarrow + \quad \Sigma F = ma_x$$

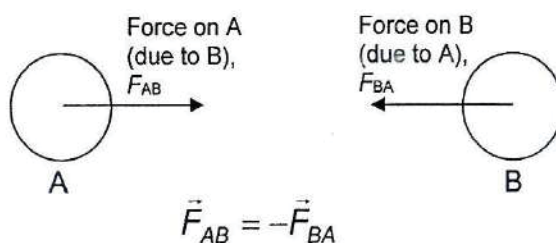
$$\begin{aligned} mg \sin \theta + (-f) &= ma \\ f &= mg \sin \theta - ma \\ &= 75(9.81 \sin 30^\circ - 2.6) \\ &= 173 \text{ N} (\leftarrow) \end{aligned}$$

$$\begin{aligned} \nearrow + \quad \Sigma F &= ma_y \\ (a=0) \quad \Sigma F &= 0 \\ N - mg \cos \theta &= 0 \\ N &= mg \cos \theta \\ &= 75(9.81) \cos 30^\circ \\ &= 637 \text{ N} (\nearrow) \end{aligned}$$

[173 N, 637 N]

3.1.6 Newton's 3rd Law**Statement of Newton's 3rd law**

If body A exerts a force on body B, then body B will exert an equal and opposite force on body A.

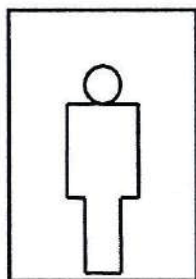


It is easy to confuse action-reaction pairs with force pairs that happen to have “equal-but-opposite” forces. Performing the following “mental checklist” may help to detect “impostors”.

For two forces to be Action-Reaction Pair, the two forces must be

- ☐ Equal in magnitude
- ☐ Opposite in direction
- ☐ Acting on different bodies
- ☐ Of the same nature. (e.g. grav-grav, friction-friction, tension-tension)

To illustrate, consider the forces when a person is standing in an elevator.



Forces acting on the person

1st action-reaction pair2nd action-reaction pair

Example 8 (Man-in-the-elevator problem)

A 70.0-kg man stands inside an elevator that travels from the ground floor to the 30th floor. A weighing scale is put under the man so that the normal contact force on his feet is measured.

The elevator is programmed to change its speed only at a rate of 7.85 m s^{-2} whenever it starts or stops. The elevator travels at a constant speed between the 5th and the 25th floor. The movement of the elevator is divided into six stages as shown. For each stage, find the reading on the weighing scale.

Elevator is going up

a) between 1st and 5th floor.

Normal contact force, N

Weight, mg

$\uparrow$

$\Sigma F = ma$

$N - mg = ma$

$N - 70.0(9.81) = 70.0(7.85)$

$N = 1240 \text{ N} \quad (\uparrow)$

Increasing upwards velocity.

Net force upwards

$\uparrow a$

b) between 5th and 25th floor.

Normal contact force, N

Weight, mg

$\uparrow$

$\Sigma F = 0$

$N - mg = 0$

$N = mg = 70.0(9.81)$

$N = 687 \text{ N} \quad (\uparrow)$

No change in velocity.

Net force = 0

c) between 25th and 30th floor.

Normal contact force, N

Weight, mg

$\downarrow$

$\Sigma F = ma$

$mg - N = ma$

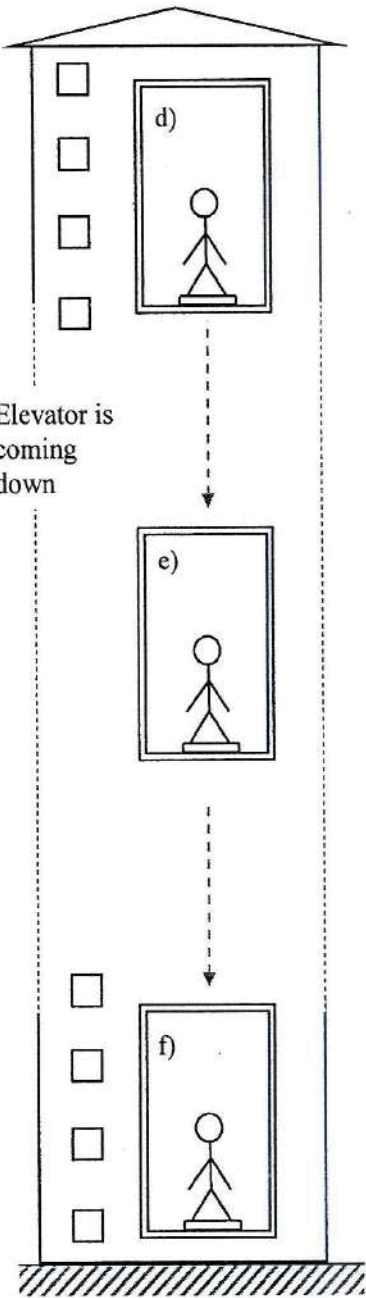
$70.0(9.81) - N = 70.0(7.85)$

$N = 137 \text{ N} \quad (\uparrow)$

Decreasing upwards velocity.

Net force downwards

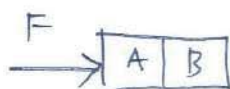
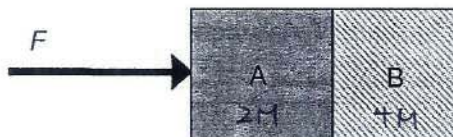
$\downarrow$

 Elevator is coming down	d) between 25th and 30th floor. 137 N
	e) between 5th and 25th floor. 687 N
	f) between 1st and 5th floor. 1240 N

Example 9

Two blocks A and B, of masses $2M$ and $4M$ respectively are pushed along a smooth horizontal surface by a force of F as shown in the diagram.

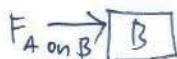
What is the magnitude of the force exerted by block A on block B during the acceleration?



$$\text{By N2L, } \Sigma F = ma$$

$$F = (2M + 4M)a$$

$$a = \frac{F}{6M}$$



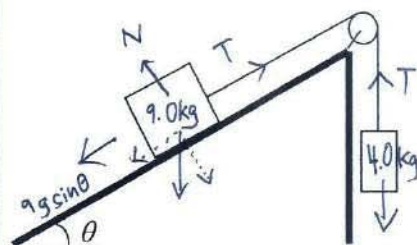
$$\text{By N2L, } F_{A \text{ on } B} = 4Ma$$

$$= 4M \left(\frac{F}{6M} \right)$$

$$= \frac{2F}{3}$$

Example 10

A 9.0 kg block is placed on a frictionless inclined plane and connected to a 4.0 kg block as shown in the diagram. If the angle θ is 35° , what would be the acceleration of the 9.0 kg block?



$$\Sigma F = ma$$

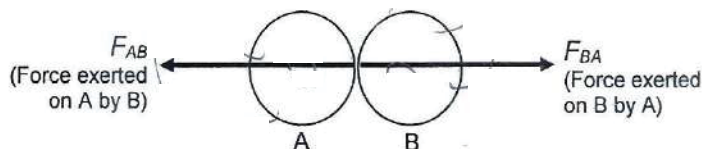
$$\text{By N2L, } 9.0g \sin 35^\circ + (4.0g) = (9 + 4)a$$

$$a = 0.88 \text{ m s}^{-2}$$

$$9.0g \sin 35^\circ \leftarrow \boxed{9.0} \xrightarrow{T} \xrightarrow{T} \boxed{4.0} \rightarrow 4.0g$$

3.2 Principle of Conservation of Linear Momentum

Momentum is an important concept because under certain circumstances, momentum is a **conserved** quantity. Consider, for example, the head-on collision of two billiard balls, A and B.



Assuming **no external forces**, the only forces experienced by the two bodies A and B during the collision will be the mutual normal contact forces, F_{AB} and F_{BA} .

By Newton's 3rd law, the two forces must be equal but opposite:

$$\vec{F}_{AB} = -\vec{F}_{BA}$$

Since the entire interaction is simultaneous and over the same duration, the impulses of the two forces must also be equal but opposite:

$$\int \vec{F}_{AB} dt = \int -\vec{F}_{BA} dt$$

$$\vec{J}_{AB} = -\vec{J}_{BA}$$

It follows that the momentum exchanged between A & B must also be equal but opposite:

$$(\Delta \vec{p})_A = -(\Delta \vec{p})_B$$

If we see the two billiard balls as making up one system, then the **total momentum of the system does not change** after the collision:

$$\begin{aligned} \Sigma \Delta \vec{p} &= (\Delta \vec{p})_A + (\Delta \vec{p})_B \\ &= (\Delta \vec{p})_A + -(\Delta \vec{p})_A \\ \Sigma \Delta \vec{p} &= 0 \end{aligned}$$

The above derivation can be extended to include any number of interacting objects in three-dimensional space, involving forces of any nature (magnetic, electric, gravitational, etc), as long as the condition of **no resultant external force** holds. The logic is surprisingly straight-forward:

$$\Sigma \vec{F} = 0 \Rightarrow \Sigma \vec{J} = 0 \Rightarrow \Sigma \Delta \vec{p} = 0$$

Thus the general statement of the principle of conservation of momentum is

Statement of the Principle of Conservation of Linear Momentum

The **total** linear momentum of a system is conserved if **no net external force** acts on the system.

Mathematically, $\Sigma m_n \vec{u}_n = \Sigma m_n \vec{v}_n$



The principle is NOT restricted to collisions!

Every year, many students get zero mark during examinations when asked to state the principle of conservation of momentum because they wrote about momentum before and after collisions and stuff.

Notes:

- By a system, we mean a set of objects that we choose.
- Some textbooks refer to the system as being *isolated*. An isolated system is one in which the only (significant) forces are those between the objects in the system. The sum of all these "internal" forces within the system will be zero because of Newton's 3rd Law.
- If there ARE external forces – then the total momentum of the system will NOT be conserved. E.g., if we take as our system a rock falling under gravity, the momentum of this system is NOT conserved because an EXTERNAL force, the gravitational pull by the Earth is acting on it. However, if we include the Earth in the system, the total momentum of rock plus Earth IS conserved.

3.2.1 Two-body System

For a **two-body system**, the principle of conservation of momentum can be mathematically simplified as follows:

$$\begin{aligned}
 (\Delta \vec{p})_A &= -(\Delta \vec{p})_B \\
 (\vec{p}_f - \vec{p}_i)_A &= -(\vec{p}_f - \vec{p}_i)_B \\
 m_A(\vec{v}_A - \vec{u}_A) &= m_B(\vec{u}_B - \vec{v}_B) \\
 m_A\vec{u}_A + m_B\vec{u}_B &= m_A\vec{v}_A + m_B\vec{v}_B
 \end{aligned}$$

Example 11

A marksman holds a rifle of mass $M = 3.00 \text{ kg}$ loosely in his hands, so as to let it recoil freely when fired. He fires a bullet of mass $m = 5.00 \text{ g}$ horizontally with a velocity of $v_B = 300 \text{ m s}^{-1}$. What is the recoil velocity v_R of the rifle?

Taking to the right as +ve,

Applying PCOLM :

$$\sum p_i = \sum p_f$$

$$0 = mv_B + M(-v_R)$$

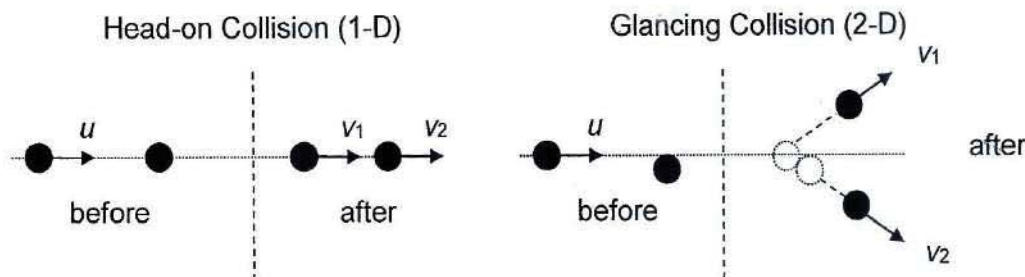
$$\Rightarrow v_R = \frac{mv_B}{M} = \frac{(0.00500)(300)}{3.00} = 0.500 \text{ m/s}$$

The recoil velocity of the rifle is 0.500 m/s towards the left.

3.2.2 Types of Collisions

A **collision** is an isolated event in which two or more colliding bodies exert **relatively strong** forces on each other for a relatively short time. Note that the two bodies do not necessarily have to be in any physical contact, as interaction may involve a force at a distance.

3.2.2.1 Head-on vs Glancing Collisions



A head-on collision is one in which the directions of motion before and after the collision are along the same line of motion.

3.2.2.2 Elastic vs Inelastic Collisions

The total kinetic energy of the colliding bodies is NOT necessarily conserved during a collision. (This is unlike the total momentum, which is ALWAYS conserved during a collision) Some kinetic energy is usually converted to heat, sound and deformation energy. How much of the original kinetic energy is conserved after the collision depends on the **elasticity**¹ of the collision.

- An **elastic** collision is one where 100% of the kinetic energy is conserved. E.g. two diamonds bouncing off each other. Note that *during* the collision, some of the kinetic energy can be temporarily be converted to potential energy.
- An **inelastic** collision is one that is NOT elastic, i.e. some or all of the KE is converted into heat, sound, deformation energy, etc.
- A **perfectly inelastic** collision is one where the 2 objects coalesce (stick together) and move with **common velocity** after the collision. E.g. two lumps of clay that don't bounce at all, but stick together after the collision. This represents the **maximum possible loss** of kinetic energy without violating the principle of conservation of energy. It does NOT necessarily mean all the KE is lost.

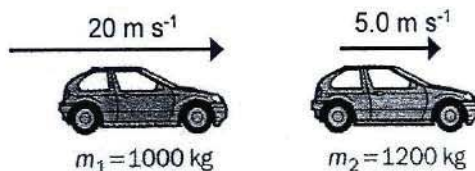
¹ Refer to Appendix B if you want to know how the elasticity is measured by the coefficient of restitution.

Example 12 (Perfectly Inelastic collision)

A car of mass 1000 kg moving at 20.0 m s^{-1} collides with a car of mass 1200 kg, moving at 5.0 m s^{-1} in the same direction.

(a) If the collision is perfectly inelastic, determine their final velocity after the collision.

(b) Calculate also the ratio of $\frac{\text{total final KE}}{\text{total initial KE}}$



a) applying **PrCLM**: $(+ \rightarrow) \sum p_i = \sum p_f$

$$1000(20) + 1200(5.0) = (1000 + 1200)V$$

$$V = 11.8 \text{ m/s}$$

b) Total K.E = $\frac{1}{2}(1000(20)^2 + \frac{1}{2}(1200)(5)^2$

$$= 215000 \text{ J}$$

$$\text{Final} = \frac{1}{2}(2200)(11.8)^2$$

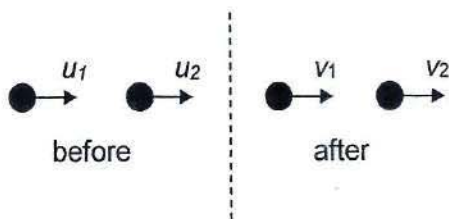
$$= 153423.71 \text{ J}$$

$$\frac{\text{Total final K.E}}{\text{Total initial K.E}} = \frac{153423.71}{215000}$$

$$= 0.71 \text{ (2sf)}$$

3.2.2.3 Relative Speed of Approach/Separation during Elastic Collision

Consider a two-body **elastic** head-on collision.



By PCOM:

$$m_1 u_1 + m_2 u_2 = m_1 v_1 + m_2 v_2$$

Rearranging, we get

$$m_1(u_1 - v_1) = m_2(v_2 - u_2) \text{ -----(1)}$$

Since KE is conserved during an elastic collision:

$$\frac{1}{2} m_1 u_1^2 + \frac{1}{2} m_2 u_2^2 = \frac{1}{2} m_1 v_1^2 + \frac{1}{2} m_2 v_2^2$$

Rearranging, we get

$$m_1(u_1^2 - v_1^2) = m_2(v_2^2 - u_2^2)$$

$\Rightarrow$

$$m_1(u_1 - v_1)(u_1 + v_1) = m_2(v_2 - u_2)(v_2 + u_2) \text{ -----(2)}$$

When (2) / (1)

$\Rightarrow$

$$(u_1 + v_1) = (v_2 + u_2)$$

Rearranging

$$(u_1 - u_2) = (v_2 - v_1),$$

or equivalently,

$$(u_1 - u_2) = -(v_1 - v_2)$$

The $(u_1 - u_2)$ term represents the speed of body 1 relative to body 2 before the collision.

The $(v_1 - v_2)$ term represents the speed of body 1 relative to body 2 after the collision.

We thus come to the following conclusion:

Statement of Relative Speed of Approach/Separation

For **elastic** collisions, relative speed of approach is equal to the relative speed of separation.

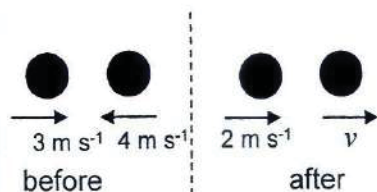
Mathematically, $u_1 - u_2 = v_2 - v_1$



This result can only be applied for (2 body) **elastic** collision. During examinations, the above statement must be **made** before it can be applied mathematically in solving problems.

Example 13

Assuming the following are head-on **elastic** collisions, solve for the unknown velocity, v in each case. Note that the masses of the objects are not (necessarily) equal.

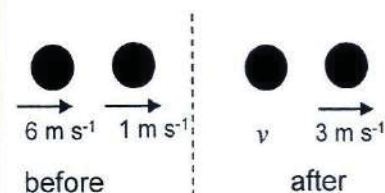


Since collision are elastic:

Rel spd of approach = Rel spd of separation

$$3 + 4 = v - 2, \quad v = 5 \text{ m/s } (\rightarrow)$$

Opp. directions: ADD



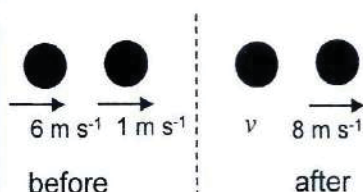
$$6 - 1 = 3 + v, \quad v = 2 \text{ m/s } (\leftarrow)$$

$$6 - 1 = 3 - v, \quad v = -2 \text{ m/s } (\leftarrow)$$

Same direction:

approaching (behind-front)

Separating (front-behind)



TIPS Solving Collision Problems

Step 0 Illustrate if no diagram is provided.

Step 1 Set up the PCOM equation, choosing an appropriate positive direction. **IF** collision is perfectly inelastic, use one variable to represent the final velocities of both bodies (i.e. make $v_1 = v_2 = v$)

Step 2 **IF** collision is elastic, set up the Relative Speed of Approach/Separation equation.

Step 4 Solve for the unknowns, e.g. mass, initial or final velocities.
(you need as many independent equations as there are unknowns to solve them uniquely)

$$\bar{u}_1 - \bar{u}_2 = \bar{v}_2 - \bar{v}_1$$

$$3 - (-4) = v_2 - 2$$

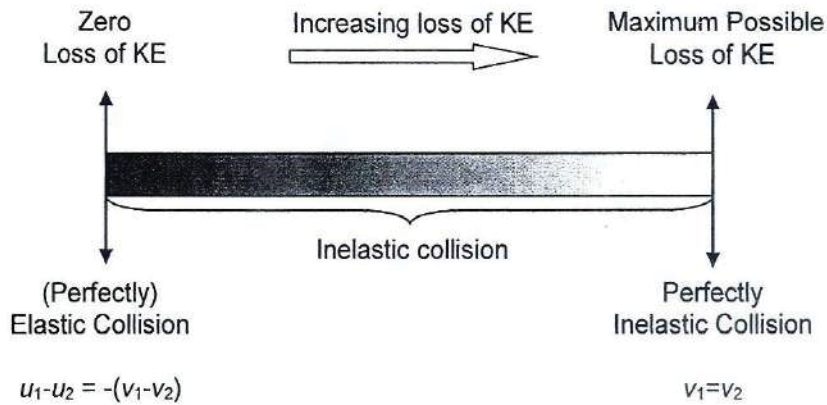
$$3 + 4 = v_2 - 2$$

$$7 = v_2 - 2$$

$$v_2 = 9$$



Total Momentum is ALWAYS conserved



Example 14 (Elastic collision)

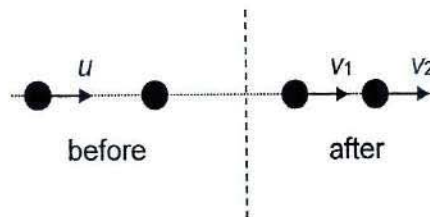
Two spheres, A and B, of mass 4.0 kg and 6.0 kg respectively, collide head-on. Their initial velocities are 5.0 m s⁻¹ and 7.0 m s⁻¹ in opposite directions, as shown in the diagram.

Assuming the collision is **elastic**, determine their final velocities.



3.2.2.4 Elastic Collision, One Body Initially at Rest (*Good to know, not stated in syllabus*)

Let's study a head-on elastic collision between two bodies, one of which is initially at rest.



By PCOM:

$$m_1 u = m_1 v_1 + m_2 v_2 \quad \text{-----(1)}$$

For elastic collision, relative speed of approach = relative speed of separation:

$$u - 0 = v_2 - v_1$$

Substitute $v_2 = u + v_1$ into (1): $m_1 u = m_1 v_1 + m_2 u + m_2 v_1$

$$v_1 = \frac{m_1 - m_2}{m_1 + m_2} u$$

Substitute $v_1 = v_2 - u$ into (1): $m_1 u = m_1 v_2 - m_1 u + m_2 v_2$

$$v_2 = \frac{2m_1}{m_1 + m_2} u$$

Case 1: $m_1 = m_2$

$$v_1 = 0, v_2 = u$$

E.g. During a billiard shot, the cue ball comes to a complete rest after hitting the target ball (of the same mass), passing all the momentum to the target ball.

Case 2: $m_1 \ll m_2$

$$v_1 \approx -u, v_2 \approx 0$$

E.g. A ping pong ball lands on the table and bounces back with the same speed as before.

Case 3: $m_1 \gg m_2$

$$v_1 \approx u, v_2 \approx 2u$$

E.g. After hitting a bowling pin, the bowling ball continues with no noticeable drop in speed, while the pin was sent flying forward at about twice the ball's speed.

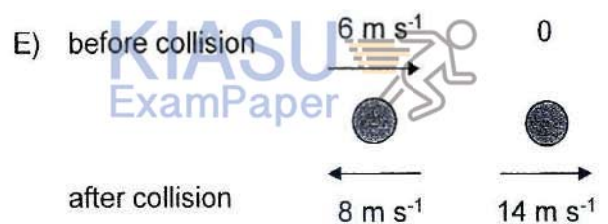
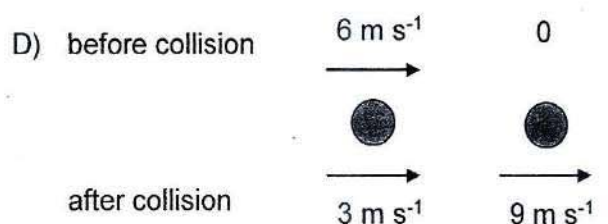
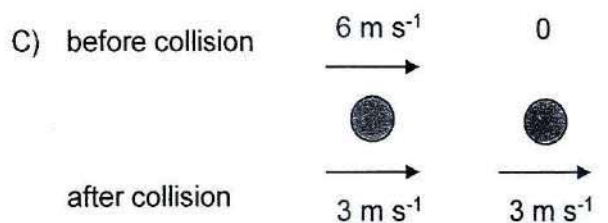
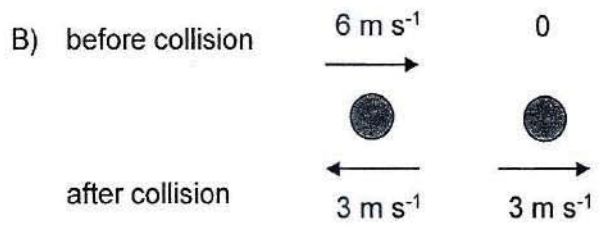
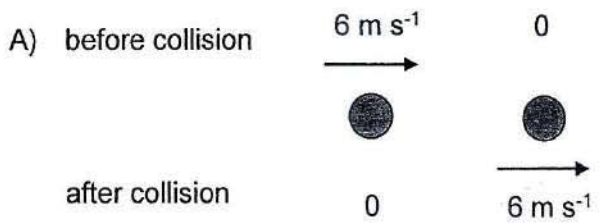


Obviously, these results are only valid for head-on elastic collision between two bodies, one initially at rest. During examinations, you are NOT allowed to quote and use these results without derivation.

Example 15 (N06/I/5 [modified])

Two spheres of unknown masses collide, with one sphere initially stationary and the other initially moving at 6 m s^{-1} . Examine the options in the diagram, then answer the following questions.

- (a) Which of the following could represent an elastic collision(s)?
- (b) Which of the following represents an elastic collision of *equal masses*?
- (c) Consider case D. Determine the ratio of m_1/m_2 .
- (d) What does case C represent?



Appendix A: Static and Kinetic Friction

Static Friction

The magnitude of static friction between two surfaces can be expressed as

$$f_s \leq \mu_s N$$

where μ_s is the coefficient of static friction between the two surfaces in question and N is the magnitude of the normal contact force between the two surfaces.

Static friction thus has a maximum magnitude called the limiting friction $f_{\max} = \mu_s N$.

Kinetic Friction (or sliding friction or dynamic friction)

The magnitude of static friction between two surfaces can be expressed as

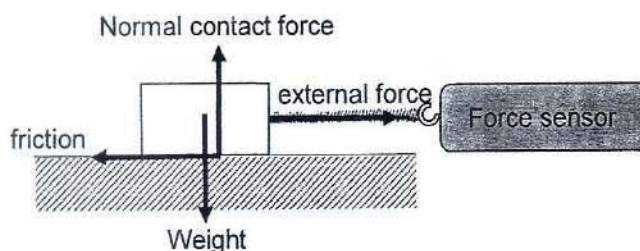
$$f_k = \mu_k N$$

where μ_k is the coefficient of kinetic friction between the two surfaces in question and N is the magnitude of the normal contact force between the two surfaces.

Unlike the static friction, kinetic friction is a constant value. It is also independent of the relative speed of motion.

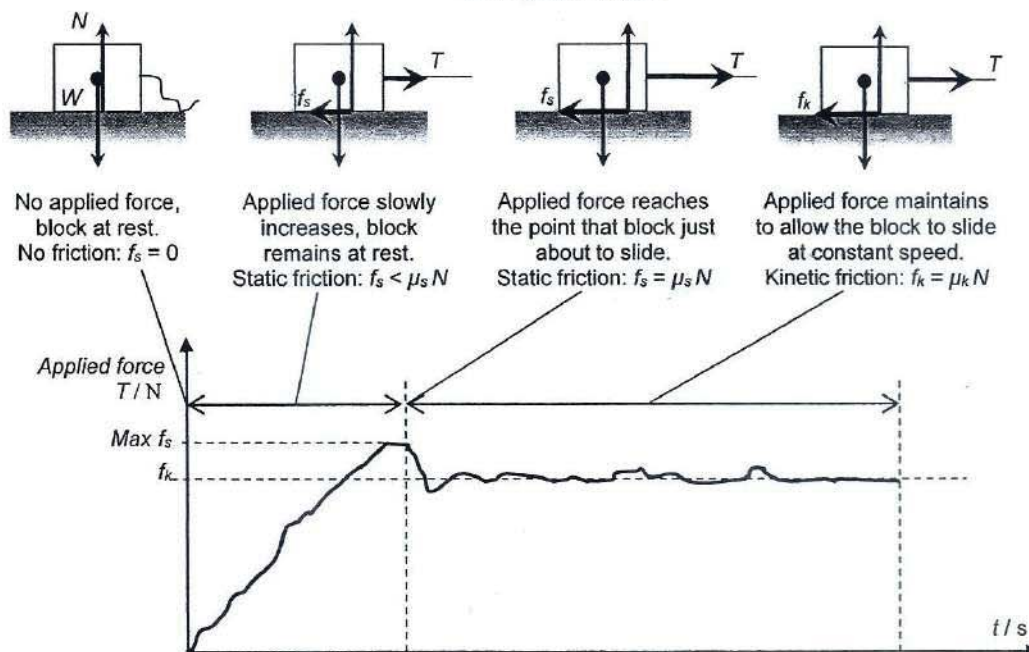
Both μ_s and μ_k are empirically found to be always less than 1 (friction is always less than normal contact force). μ_k is found to be always less than μ_s (kinetic friction is always smaller than the limiting static friction).

To better understand the two, consider the results of the following experiment.



An external force is applied on the block and is increased very gradually until the block starts moving and continues to move at a constant speed. Since the block is always either at static equilibrium or constant speed motion, the magnitude of frictional force can be inferred from the applied force as measured by the force sensor.

The graph of the applied force against time is shown below in relation to the magnitudes of friction, both static and kinetic:



Appendix B: Centre of Mass

Centre of Mass (Good to know, Not stated in syllabus)

If a mechanical system can be reduced to a set of particles of known mass and spatial separation, then the centre of mass for the original rigid body can be calculated by summing the moments for each particle about any line we choose, and then calculating the distance from the line for a particle equal in mass to the total mass such that the moment of the particle of total mass is the same as the sum of the moments for all the particles. The sum of moments must account for a specified rotational sense with respect to the chosen line.

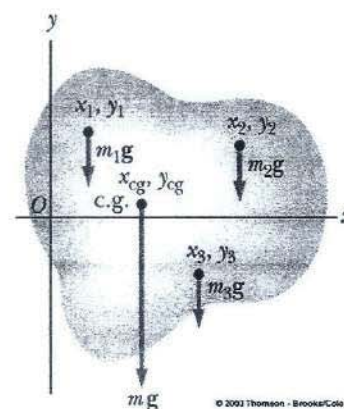
Suppose we have several particles, with masses m_1 , m_2 and so on. Let the coordinates of m_1 be (x_1, y_1) and those of m_2 be (x_2, y_2) and so on.

The centre of mass of the system is the point (x_{cm}, y_{cm}) where

$$x_{cm} = \frac{m_1 x_1 + m_2 x_2 + m_3 x_3 + \dots}{m_1 + m_2 + m_3 + \dots} = \frac{\sum m_i x_i}{M}$$

$$y_{cm} = \frac{m_1 y_1 + m_2 y_2 + m_3 y_3 + \dots}{m_1 + m_2 + m_3 + \dots} = \frac{\sum m_i y_i}{M}$$

M denotes the total mass $m_1 + m_2 + m_3 + \dots$ of the particles.



Appendix C: Coefficient of Restitution

The elasticity of a collision is quantified by a number called the coefficient of restitution, ε , defined as the ratio of the speed of separation to the speed of approach:

$$\varepsilon = \frac{|\vec{v}_2 - \vec{v}_1|}{|\vec{u}_2 - \vec{u}_1|}$$

A perfectly elastic collision has a coefficient of restitution of 1. Example: two diamonds bouncing off each other. A perfectly plastic, or inelastic, collision has $\varepsilon = 0$. Example: two lumps of clay that don't bounce at all, but stick together. So the coefficient of restitution will always be between zero and one.

ε is determined by several factors like the material of the colliding bodies, and *assumed* to be independent of speed.

Appendix D: Newton's Law in Car Safety

Using Newton's laws, discuss how each of the following features in vehicles increases the safety of its passengers:

- a) Seat Belts
- b) Head restraints
- c) Air Bags

a) Seat belt exerts a retardation force on the passenger, without which he would have continued in his state of uniform motion in a straight line into the windscreen. (N1L)

(The vehicle is able to decelerate rapidly because the braking system leads to a frictional force between the road surface and the tyres. Unfortunately, this frictional force acts on the vehicle, not the passenger. Without the seat belt, the only retardation force would be the frictional force between the seat and the passenger's bum. But this force is limited. If the vehicle decelerates too fast, the frictional force is not large enough to decelerate the passenger at the same rate as the vehicle)

(Be very clear that when passengers crash into the wind screen, it was due to their inertia. It was not due to some mysterious "forward force")

b) If the car is hit from the rear, the head restraint provides a forward force on the passenger's head so that the head is accelerated at the same rate as the rest of his body, so that he does not suffer from whiplash (some kind of neck injury).

c) Airbags exerts a relatively small force over a relatively long duration of time, safely bringing the passenger to rest by reducing the passenger's momentum at a relatively slower rate. (N2L)

Topic 4: Forces

H2 Physics C1 2016

H2 Physics Syllabus 9749

Topic 4: Forces

Content

- Types of force
- Equilibrium of forces
- Centre of gravity
- Turning effects of forces

Learning Outcomes

Candidates should be able to:

- (a) recall and apply Hooke's law ($F = kx$, where k is the force constant) to new situations or to solve related problems.
- (b) describe the forces on a mass, charge and current-carrying conductor in gravitational, electric and magnetic fields, as appropriate. (*covered later in relevant chapters.*)
- (c) show a qualitative understanding of normal contact forces, frictional forces and viscous forces including air resistance. (No treatment of the coefficients of friction and viscosity is required.) (*partly covered in Kinematics.*)
- (d) show an understanding that the weight of a body may be taken as acting at a single point known as its centre of gravity. (*covered in Dynamics.*)
- (e) define and apply the moment of a force and the torque of a couple.
- (f) show an understanding that a couple is a pair of forces which tends to produce rotation only.
- (g) apply the principle of moments to new situations or to solve related problems.
- (h) show an understanding that, when there is no resultant force and no resultant torque, a system is in equilibrium.
- (i) use a vector triangle to represent forces in equilibrium.
- (j) derive, from the definitions of pressure and density, the equation $p = \rho gh$.
- (k) solve problems using the equation $p = \rho gh$.
- (l) show an understanding of the origin of the upthrust acting on a body in a fluid.
- (m) state that upthrust is equal to the weight of the fluid displaced by a submerged or floating object.
- (n) calculate the upthrust in terms of the weight of the displaced fluid.
- (o) recall and apply the principle that, for an object floating in equilibrium, the upthrust is equal to the weight of the object to new situations or to solve related problems.

4.1 Hooke's Law

In 1676, Robert Hooke stated an empirical law bearing his name.

Hooke's Law states that the magnitude of force F exerted by a spring on a body attached to the spring is proportional to the extension x of the spring from equilibrium provided the proportional limit of the spring is not exceeded.

$$F = kx$$

k = constant of proportionality (called force constant or spring constant or stiffness of spring)

Example 1 (N94/I/23)

A spring, obeying Hooke's Law, has an unstretched length of 50 mm and a spring constant of 400 N m^{-1} . What is the tension in the spring when its overall length is 70 mm?

$$\begin{aligned} F &= 400 \times (70 - 50) \times 10^{-3} \\ &= 8.0 \text{ N} \end{aligned}$$

4.2 Upthrust or Buoyant Force

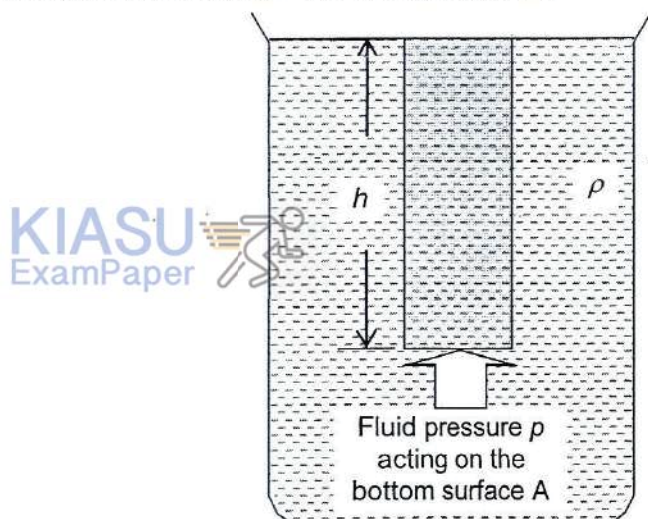
Before we discuss upthrust, we must first familiarize ourselves with pressure exerted by a fluid.

4.2.1 Pressure due to Fluid

Pressure due to fluid is the force acting per unit area by the fluid.

Examples are water pressure (hydrostatic pressure) and atmospheric pressure. Water pressure exists in a vertical column of water because water at any depth has to support the weight of column of water above. This pressure **increases with depth** because the weight of column of water to be supported increases with depth. In fact any fluid in a gravitational field including the atmosphere will exert a pressure that increases with depth in the fluid.

4.2.2 Derivation of formula for Pressure due to Fluid



Consider the forces acting on a column of fluid as shown in the diagram.

Let ρ (Greek letter rho) = density of the fluid,
 A = surface area of the bottom of the column,
 h = depth from the surface of the fluid,
 m = mass of the fluid column

By Newton's Laws, the net force acting on the column of fluid is zero because the column is stationary (it is part of a uniform incompressible fluid which is also stationary). Thus, the fluid force F acting on the bottom surface is equal to the weight of the column of fluid, mg .

$$F = mg$$

Since the fluid force $F = pA$

mass of fluid $m = \rho Ah$

(Recall that density ρ is mass m per unit volume V , hence $m = \rho V = \rho Ah$)

Hence,

$$pA = \rho Ahg$$

$$p = \rho hg$$

The pressure p due to fluid is given by

where

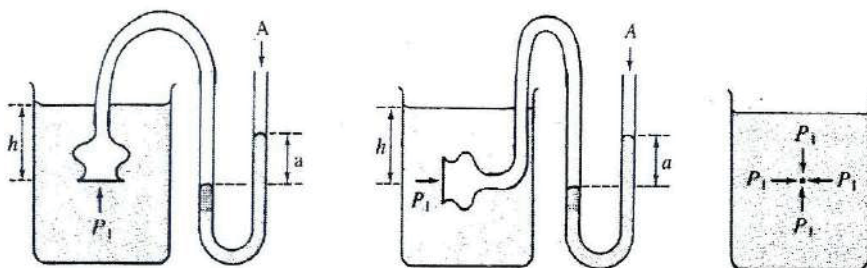
$$p = \rho hg$$

ρ = density of fluid

g = gravitational acceleration

h = vertical depth from the surface of the fluid.

Because a fluid can flow in all directions, the pressure due to the fluid acts in **all directions**, as depicted by the diagram below.

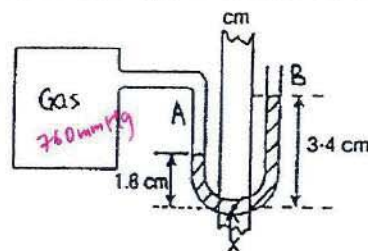


If the surface of the fluid experiences an external pressure p_o (e.g. atmospheric pressure), the total pressure in a fluid is the sum of the pressure due to the fluid and the external pressure p_o acting on the surface of the fluid.

$$p = \rho gh + p_o$$

Example 2

The diagram on the right shows a mercury manometer connected to a gas container. The atmospheric pressure is 760 mm Hg and the point X is at the bottom of the U-tube.



- What is the pressure of the point X in mm Hg?
- What is the pressure of the gas in mm Hg?
- If the manometer is disconnected from the gas container, what will be the vertical heights of the mercury levels from the point X in the limbs A and B?

$$a) \text{ At } X, P = P_0 + 3.4 \text{ cm Hg}$$

$$= 794 \text{ mmHg}$$

$$P = \frac{794}{760} P_0 = 1.05 P_0$$

$$= 1.06 \times 10^5 \text{ Pa}$$

$$(P_0 = 1.01 \times 10^5 \text{ Pa})$$

$$x \text{ mmHg} \rightarrow P_g$$

$$\frac{x}{760} \times 1.01 \times 10^5 \text{ Pa}$$

$$b) P_{\text{gas}} = P_0 + 1.6 \text{ cmHg}$$

$$= 776 \text{ mmHg}$$

$$P = \dots = 1.02 P_0$$

$$= 1.03 \times 10^5 \text{ Pa}$$

$$c) \frac{1.6}{2} = 0.8 \text{ cm}$$

$$1.8 + 0.8 = 2.6 \text{ cm}$$

$$3.4 - 0.8 = 2.6 \text{ cm}$$

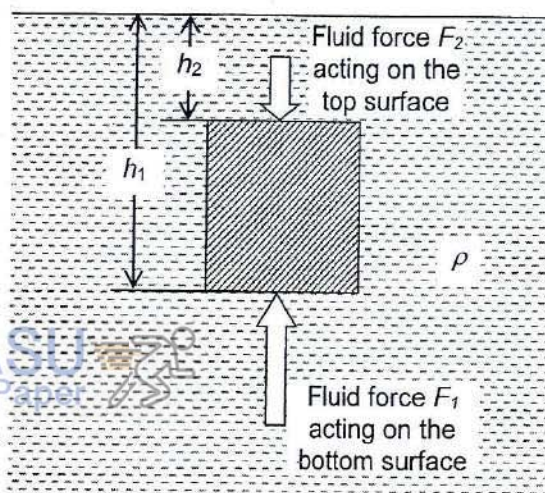
Both ~~sides~~ sides
settle at same
final height.

4.2.3 Upthrust

The variation of pressure with depth implies that the forces exerted by a fluid on a body fully or partially immersed in the fluid vary with the vertical height of the body in the fluid. This results in a net upward force on the body by the fluid, which is called an upthrust.

Definition of Upthrust

The **net upward** force exerted by a fluid on a body **fully or partially** submerged in the fluid.



By considering the fluid forces acting on a cube fully immersed in the fluid as shown in the diagram, we can generalise the result to find the upthrust experienced by any body immersed in the fluid.

Let ρ = density of the fluid,
 A = surface area of a face of the cube,
 $\Delta h = h_1 - h_2$ = length of a side of the cube.

The fluid forces acting on the side faces of the cube are equal in magnitude and in opposite directions, thus resulting in **zero net horizontal force**.

The fluid force F_1 acting on the **bottom** surface is $F_1 = \rho g h_1 A$

The fluid force F_2 acting on the **top** surface is $F_2 = \rho g h_2 A$

Thus, the **net vertical force** is $F_1 - F_2 = \rho g h_1 A - \rho g h_2 A$

$$= \rho g A (h_1 - h_2)$$

$$= \rho g A \Delta h$$

$$= \rho g V, \text{ where volume of the cube } V = A \Delta h$$

$$= \rho V g$$

$$= (\text{mass of fluid displaced}) \times g$$

$$= \text{weight of fluid displaced.}$$

Archimedes' Principle

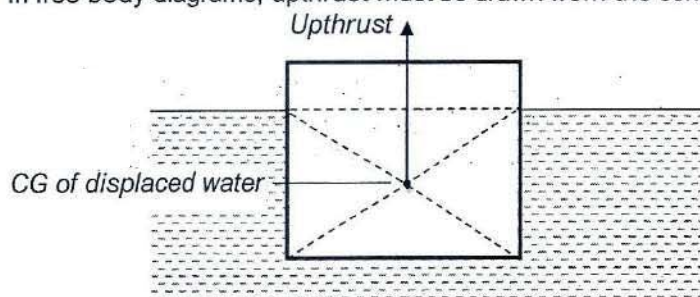
For a body submerged in a fluid, the upthrust acting on the body is equal to the **weight of fluid displaced**.

$$J = \rho V g$$

where U = upthrust, ρ = density of fluid, V = volume of fluid displaced by the body, g = gravitational acceleration

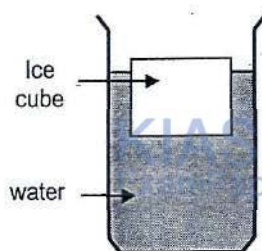


In free body diagrams, upthrust must be drawn from the centre of gravity of the fluid displaced.



Principle of Flotation

For a body floating in a fluid, the weight of the fluid displaced by the body is equal to the weight of the body.



For example, since the ice cube floats on water, the weight of the water displaced by the ice cube must be equal to the weight of the ice cube.

When the ice cube melts, what happens to the water level?

Example 3

Icebergs are often a danger to ships since most of an ice-berg's volume lies beneath the surface of the water. One can only see the "tip of the iceberg". Find the percentage of the iceberg that is submerged.

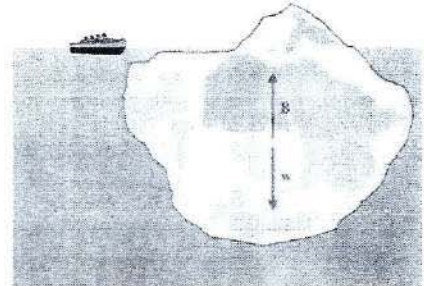
(density of ice, $\rho_i = 0.917 \text{ g cm}^{-3}$, density of sea-water, $\rho_w = 1.025 \text{ g cm}^{-3}$)

Iceberg floating

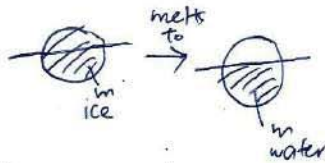
$\Rightarrow$ Upthrust on it = weight of it

$$\Rightarrow V_w \rho_w g = V_i \rho_i g$$

$$\Rightarrow \frac{V_w}{V_i} = \frac{\rho_i}{\rho_w} = \dots = 0.895 \begin{cases} 10.5\% \text{ above water} \\ 89.5\% \text{ below water} \end{cases}$$



Note: $\underbrace{V_w \rho_w}_{m_w} = \underbrace{V_i \rho_i}_{m_i}$

**4.3 Translational Equilibrium**

By Newton's Laws of Motion, when the net external force acting on a body is zero, the acceleration will be zero. The centre of mass of this body will be observed to be either stationary or moving at constant velocity. This body is then said to be in **translational equilibrium**.

Condition for Translational Equilibrium:

The net external force acting on a body is zero, i.e. $\Sigma F_{ext} = 0 \text{ N}$.

Example 4 (Resolution of forces into component method)

Four forces act on a particle as shown below such that it is in equilibrium. Find the magnitude and direction of the unknown force F that maintains the equilibrium.

$$\text{In eqn} \Rightarrow \Sigma \vec{F} = \vec{F}_1 + \vec{F}_2 + \vec{F}_3 + \vec{F}_4 = 0 \Rightarrow \Sigma F_x = 0, \Sigma F_y = 0$$

$$(\rightarrow) \Sigma F_x = F_{1x} + F_{2x} + F_{3x} + F_{4x} = 0$$

$$= 0 + (8.5 \cos 45^\circ) + (-7.0 \cos 60^\circ) + F_{4x} = 0$$

$$\Rightarrow F_{4x} = -2.5104 \text{ N}$$

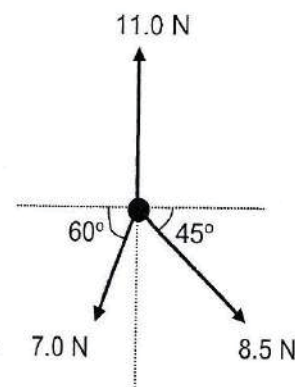
$$(+\uparrow) \Sigma F_y = F_{1y} + F_{2y} + F_{3y} + F_{4y} = 0$$

$$= (11.0) + (-8.5 \sin 45^\circ) + (-7.0 \sin 60^\circ) + F_{4y} = 0$$

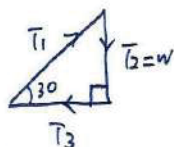
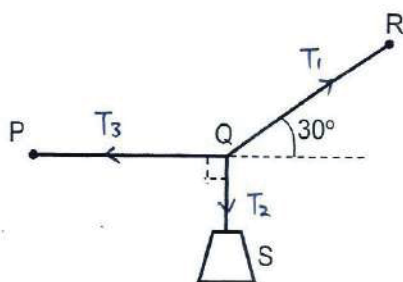
$$\Rightarrow F_{4y} = 1.0726 \text{ N}$$

$$|\vec{F}_4| = \sqrt{F_{4x}^2 + F_{4y}^2} = \dots 2.73 \text{ N}$$

$$\theta = \tan^{-1}\left(\frac{1.0726}{-2.5104}\right) = 22.1^\circ \quad (23.1^\circ \text{ clockwise of } -x \text{ direction})$$

**Example 5 (J85/I/2) (vector triangle method)**

In the diagram below, a body S of weight W hangs vertically by a thread tied at Q to the string PQR . If the system is in equilibrium, what is the tension in the section PQ ?



$$\tan 30^\circ = \frac{T_2}{T_3}$$

$$T_3 = \frac{T_2}{\tan 30^\circ} = \frac{W}{\tan 30^\circ}$$

4.4 Moment of a Force

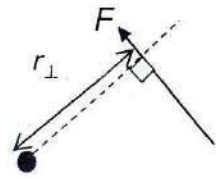
Besides translational motion, objects also exhibit rotational motion. Forces applied on a rigid body¹ can also cause or change its rotational motion, other than translational motion, that we are quite familiar with. The quantitative measure of the extent to which a force can cause or change the rotational motion of a rigid body is called the **moment of the force**, τ (Greek letter tau). The **moment of a force** is also called the **torque**.

¹ In physics, we sometimes idealise objects as **particles**. This means we ignore the size of the object and treat it like a point-like body. A **particle** has no spatial dimensions, but can have mass. It follows that particles are capable of translational motion only. Obviously we cannot treat objects as particles if we were to analyse their rotational motion. Neglecting that a real-world body may deform (e.g. stretch, bend, twist, etc) we assume that the body has a perfectly definite and unchanging shape and size. We call this idealized model a **rigid body**.

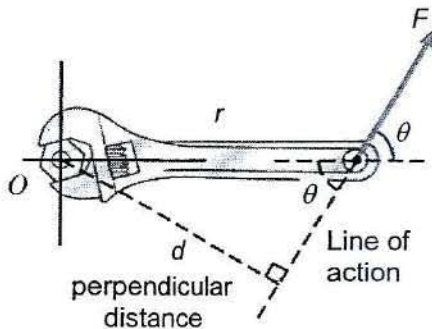
Definition of Moment of a Force

The **moment of a force** about a point is the product of the (magnitude of the) force and the perpendicular distance of the line of action of the force to the point.

$$\tau = F(r_{\perp})$$

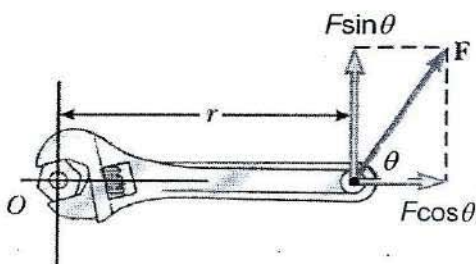


To illustrate the concept of perpendicular distance and line of action, let us consider the example of a spanner being used to loosen a bolt. A force F is applied at an acute angle θ with the length of the spanner and at a distance r away from the pivot O as shown in the diagram below.

**Method 1: Finding the perpendicular distance**

The line of action of a force is the line along which the force acts. The perpendicular distance d is the length of the line that joins the pivot O to the line of action perpendicularly. Using simple trigonometry, it can be deduced that $d = r \sin \theta$. The torque of the force F about the pivot O can then be calculated as

$$\tau = (F)(r_{\perp}) = (F)(r \sin \theta) = Fr \sin \theta$$

**Method 2: Finding the perpendicular component of force**

Another method is to resolve the force into two components: perpendicular and parallel to the spanner length. Only the perpendicular component $F \sin \theta$ produces the turning effect. The line of action of the parallel component $F \cos \theta$ passes through the pivot and produces no turning effect. The torque of the force F about the pivot O is defined as

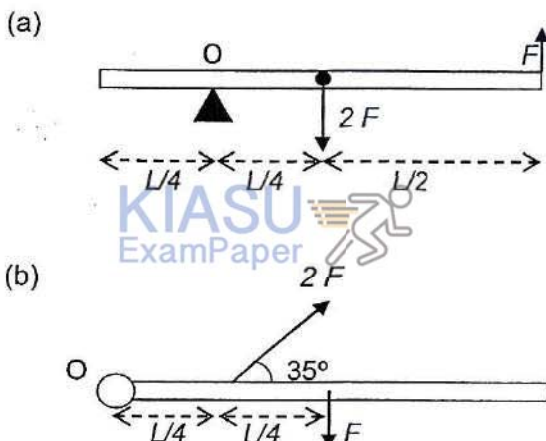
$$\tau = (r)(F_{\perp}) = (r)(F \sin \theta) = Fr \sin \theta.$$

Since both approaches lead to the formulae $\tau = Fr \sin \theta$, one could treat the equation as either

1. magnitude of force $\times$ perpendicular distance (between the pivot point and line of action),
- or
2. distance (from pivot point to point of application of force) $\times$ perpendicular component of the force.

Example 6

Find the resultant moment about point O .



Taking clockwise as +ve,

$$2F \times \frac{L}{4} = \frac{FL}{2}$$

$$F \times \frac{3L}{4} = \frac{3FL}{4}$$

$$2F \sin 35^\circ \times \frac{L}{4} =$$

$$F' \times \frac{L}{2} = \frac{FL}{2}$$

$$0.213 FL \text{ clockwise}$$

$$\frac{3FL}{4} - \frac{FL}{2} = \frac{FL}{4} \text{ anticlockwise}$$

4.5 Rotational Equilibrium and Principle of Moments

Condition for Rotational Equilibrium:

The net moment on the body *about any point* is zero. $\Sigma \tau = 0 \text{ Nm}$.

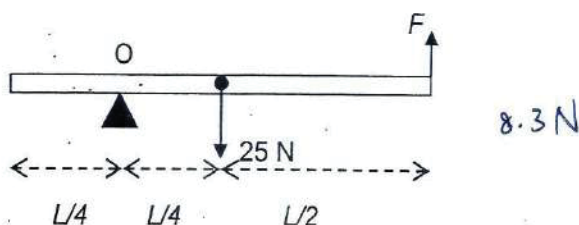
When all lines of action of the forces are in the same plane (co-planar forces), the condition for rotational equilibrium is simplified to the principle of moments.

Principle of moments states that for a body to be in rotational equilibrium, sum of clockwise moments about any point equals sum of anticlockwise moments about that same point.

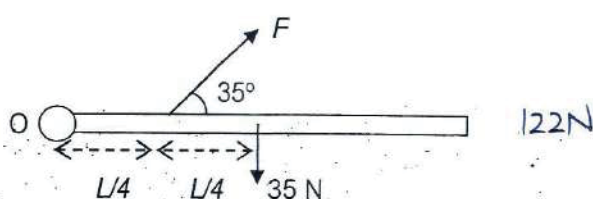
Example 7

Find the unknown force F in each of the following situations required to maintain rotational equilibrium.

(a)



(b)



4.6 Static Equilibrium

Equilibrium is a concept to describe the state of a body that is at rest or at constant velocity. Static equilibrium is when the body is at rest.

Conditions for (Static) Equilibrium:

The net external force acting on the body is zero. $\Sigma F_{\text{ext}} = 0 \text{ N}$.

The net torque on the body *about any point* is zero. $\Sigma \tau = 0 \text{ Nm}$.

TIPS

Solving Static Equilibrium problems

1. Isolate the body of interest.
2. Draw the free-body diagram and identify **all** forces acting on the body.
3. Set up the torque equation (condition for rotational equilibrium)
 - Choose a "good" pivot point.
 - A "good" pivot point is a point which lies on the lines of action of one or more of the unknown forces, thereby reducing the number of unknown terms in the equation.
 - Apply the condition $\Sigma \tau = 0 \text{ Nm}$.

Summation of clockwise moments = Summation of anti-clockwise moments

4. Set up the two force equations (x and y components) (conditions for translational equilibrium)
 - Choose two perpendicular directions x and y .
 - Resolve all forces into x and y components.
 - Apply the condition $\Sigma F_x = 0 \text{ N}$.

Summation of leftward forces = Summation of rightward forces

 - Apply the condition $\Sigma F_y = 0 \text{ N}$.

Summation of upward forces = Summation of downward forces
5. Solve the equations to find the unknown forces.

Example 8 (J95/I/5)

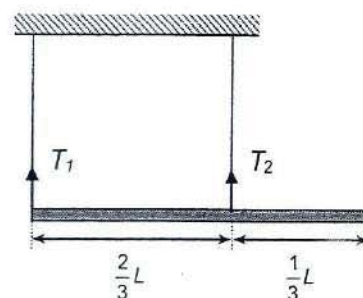
A heavy uniform beam of length L and mass M is supported by two vertical cords as shown in the diagram.

What is the ratio $\frac{T_1}{T_2}$ of the tensions in these cords?

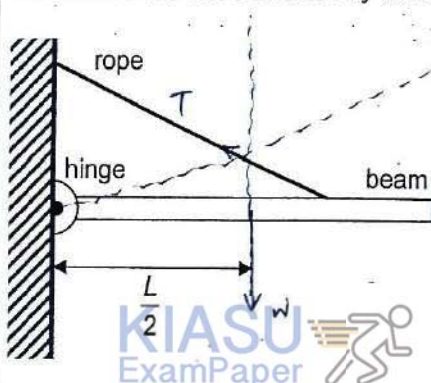
Taking moments about point O,

$$(T_1)\left(\frac{1}{2}L\right) = (T_2)\left(\frac{2}{3}L - \frac{1}{2}L\right)$$

$$\therefore \frac{T_1}{T_2} = \frac{\left(\frac{1}{2}L\right)}{\left(\frac{1}{6}L\right)} = \frac{1}{3}$$

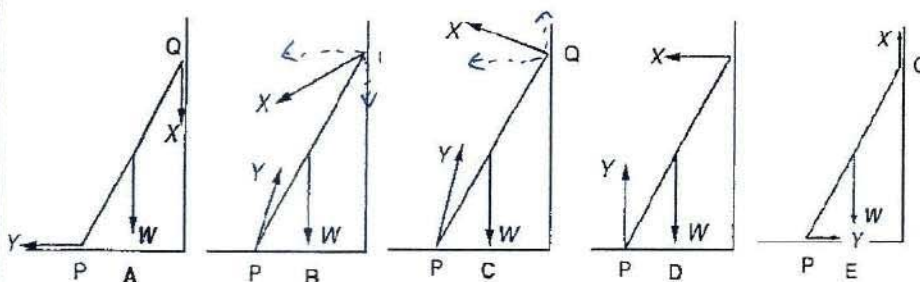
**4.7 Three-Force Systems****Example 9**

A uniform beam of length L is hinged to a wall and is supported by a rope as shown. Show the direction of the force exerted by the hinge (attached to the wall) on the beam.



Example 10 (N84/II/1)

A ladder PQ, resting on a rough floor and leaning against a rough wall, is on the point of slipping. It is of weight W and the contact forces exerted on the ladder by the wall and floor are X and Y respectively. Which one of the following diagrams correctly shows the directions of these forces?



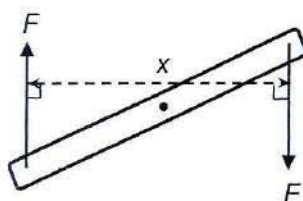
C

4.8 Couple**Definition of a Couple**

A pair of forces, which are equal in magnitude but opposite in direction, whose lines of action do not coincide.

Couples produce **only rotation** but no translation.

An example of a couple is shown in the figure below.



Taking moments about the centre of mass,

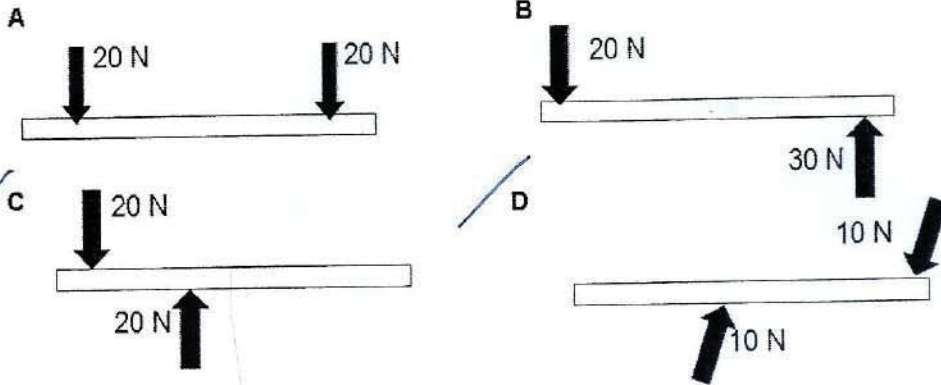
$$\begin{aligned} \text{Torque due to the couple} &= \left(\frac{1}{2}x\right)F + \left(\frac{1}{2}x\right)F \\ &= xF \text{ (in the clockwise direction)} \end{aligned}$$

In fact, taking moments about **any point** would produce the same result.

Torque of a couple is the product of the perpendicular distance between the lines of action of the forces and the magnitude of one of the forces.

Example 11

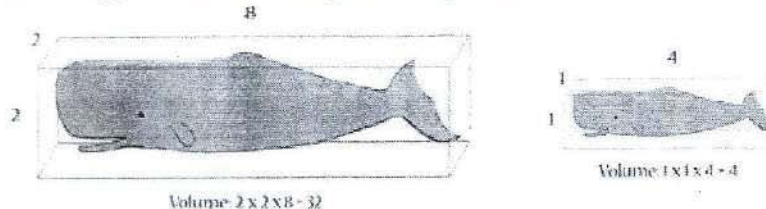
The following diagrams show forces acting on uniform rulers placed on a horizontal smooth tabletop. Which of the following diagram/s does not show a couple? Assume all forces are parallel to each other.



Appendix

Why are whales so huge?

On land, an animal is a prisoner of its own weight. The larger an animal gets, the more difficult it becomes to support itself. For example, imagine two whales, one twice the size of the other. As the measurements show, the larger whale is twice as long, twice as wide and twice as thick as the smaller whale. As you may have expected, the mass of the bigger whale is actually about eight times that of the smaller one, instead of twice.



On land, this extra weight can pose a big problem. If an elephant doubles in size, its bones cannot just double in strength – they need to get eight times as strong to support its own weight. As land animals get larger and larger, supporting body weight becomes a bigger and bigger concern. This is why the biggest land animals – such as elephants and hippos – have such thick bones.

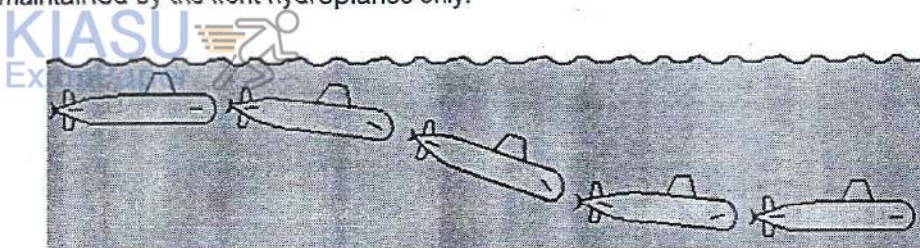
In water, however, weight is not much of a problem. Objects submerged in water are supported by the buoyant force (upthrust) due to water. The bulk of a whale's mass is made up of water (can you recall the percentage by mass of water in human?) and has similar average density to water. As a whale gets bigger, its weight gets large, and the buoyant force exerted on it increases too. Hence, a whale experiences some kind of weightlessness in water. This allows them to grow to enormous sizes without the need to develop extremely strong bones.

Buoyancy of a submarine

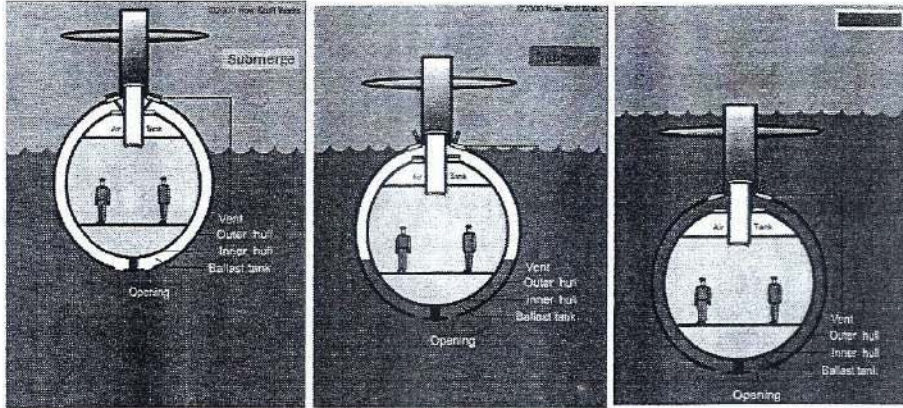
A submarine or a ship can float because the weight of water that it displaces is equal to the weight of the ship. This displacement of water creates an upward force called the buoyant force (upthrust) and acts opposite to gravity, which would pull the ship down. Unlike a ship, a submarine can control its buoyancy, thus allowing it to sink and surface at will.

To control its buoyancy, the submarine has ballast tanks and auxiliary, or trim tanks, that can be filled with water or air. When the submarine is on the surface, the ballast tanks are filled with air and the submarine's overall density is less than that of the surrounding water. As the submarine dives, the ballast tanks are flooded with water and the air in the ballast tanks is vented from the submarine until its overall density is greater than the surrounding water and the submarine begins to sink. A supply of compressed air is maintained aboard the submarine in air flasks for life support and for use with the ballast tanks.

In addition, the submarine has movable sets of short "wings" called hydroplanes on the stern (back) and bow (front) that help to control the angle of the dive. At the start of the dive the back hydroplanes are rotated upwards so that the stern of the boat is forced upwards. The front hydroplanes are rotated downwards, thus forcing the bow of the boat down. During the dive the back hydroplanes are moved to the neutral position and the dive angle is controlled with the front hydroplanes only. Close to the required depth, the back hydroplanes are rotated down and the front hydroplanes up to level off the boat. At slow speeds the depth of the boat is maintained by the front hydroplanes only.



To keep the submarine level at any set depth, the submarine maintains a balance of air and water in the trim tanks so that its overall density is equal to the surrounding water. When the submarine reaches its cruising depth, the hydroplanes are leveled so that the submarine travels level through the water. The submarine can steer in the water by using the tail rudder to turn starboard (right) or port (left) and the hydroplanes to control the front-back angle of the submarine. In addition, some submarines are equipped with a retractable secondary propulsion motor that can swivel 360 degrees.

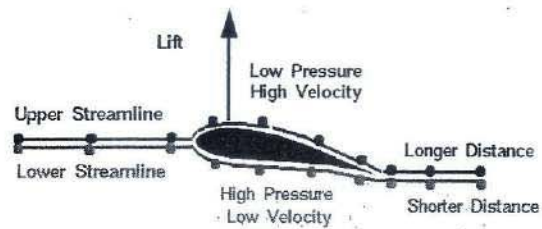


When the submarine surfaces, compressed air flows from the air flasks into the ballast tanks and the water is forced out of the submarine until its overall density is less than the surrounding water and the submarine rises. The hydroplanes are angled so that the submarine is angled upward. In an emergency, the ballast tanks can be filled quickly with high-pressure air to take the submarine to the surface very rapidly.

What keeps airplanes in the air?

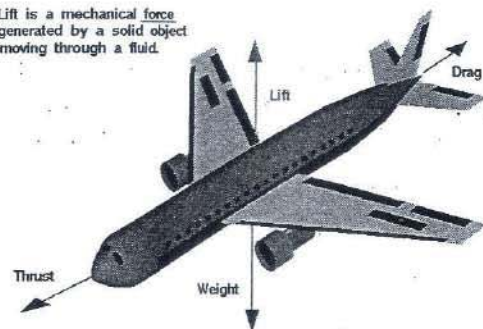
Bernoulli's principle states that the pressure of a moving gas decreases as its velocity increases. When air flows over a wing having a curved upper surface and a flat lower surface, the flow is faster across the curved surface than across the plane one; thus a greater pressure is exerted in the upward direction. The overall effect is the presence of a net upward force acting on the wing, known as lift.

The magnitude of the lift depends on the difference between the pressure on the upper side and lower side of the wing. The faster the forward motion, the greater will be the difference in the pressure. The object needs to travel at high speeds for the lift force to be significant. The wings of aeroplanes are designed to maximise the lift in order to support its weight and hence able to stay in flight.



Cross sectional view of the wing of a plane flying to the left.

Lift is a mechanical force generated by a solid object moving through a fluid.



Wings on Race Cars

In the early days of race car design, the principal aerodynamic goal was to have a design that minimized drag in order to maximize top speed. Over the years, however, as race cars started to become faster and more powerful, the need for traction, or adhesive friction between the tires and road surface, during high-speed corners began to become a pressing issue.

The easiest and most effective solution to this problem came in the form of wings or aerofoils similar to those found on airplanes, which, however, would be now designed to generate, instead of lift, "downforce," i.e. effectively negative lift. These wings would be mounted to the chassis of a race car to transfer the force generated to the chassis itself and through this to the car axles and wheels. This increases the downward pressure in the contact area between the tire and the road surface, thus also increasing the adhesive friction between these two surfaces. A typical race car wing arrangement is illustrated in Fig. 1 under the labels "A" and "B", "A" being the front wing integrated into the body of the car, and "B" the rear wing element.

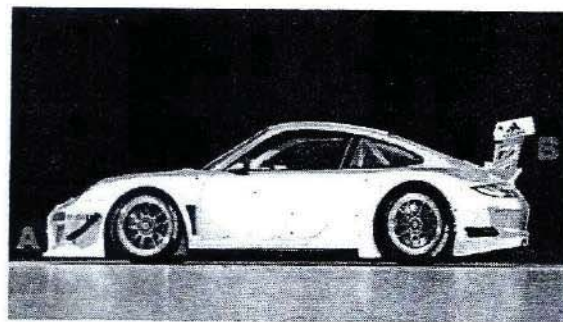


Figure 1. A Porsche 911 GT3 Racecar with front and rear wings.

It is worth noting that due to the physical constraints on the dimensions of the wings in a car, usable downforce is only generated when the race car is traveling at high speed. Due to the importance of wings in providing better traction, cornering and overall performance for race cars, much attention and research is focused on their design and implementation on the car body.

Topic 5: Work, Energy and Power

H2 Physics C1 2016

H2 Physics Syllabus 9749

Topic 5: Work, Energy and Power

Content

- Work
- Energy conversion and conservation
- Potential energy and kinetic energy
- Power

Learning Outcomes

Candidates should be able to:

- (a) show an understanding of the concept of work in terms of the product of a force and displacement in the direction of the force.
- (b) calculate the work done in a number of situations including the work done by a gas which is expanding against a constant external pressure: $W = p\Delta V$. (to be covered later in chapter on Thermal Physics.)
- (c) give examples of energy in different forms, its conversion and conservation, and apply the principle of energy conservation.
- (d) show an appreciation for the implications of energy losses in practical devices and use the concept of efficiency to solve problems.
- (e) derive, from the equations for uniformly accelerated motion in a straight line, the equation $E_k = \frac{1}{2}mv^2$.
- (f) recall and use the equation $E_k = \frac{1}{2}mv^2$.
- (g) distinguish between gravitational potential energy, electric potential energy and elastic potential energy.
- (h) deduce that the elastic potential energy in a deformed material is related to the area under the force-extension graph.
- (i) show an understanding of and use the relationship between force and potential energy in a uniform field to solve problems. (in Appendix to be covered in later chapters.)
- (j) derive, from the definition of work done by a force, the equation $E_p = mgh$ for gravitational potential energy changes near the Earth's surface.
- (k) recall and use the equation $E_p = mgh$ for gravitational potential energy changes near the Earth's surface.
- (l) define power as work done per unit time and derive power as the product of a force and velocity in the direction of the force.

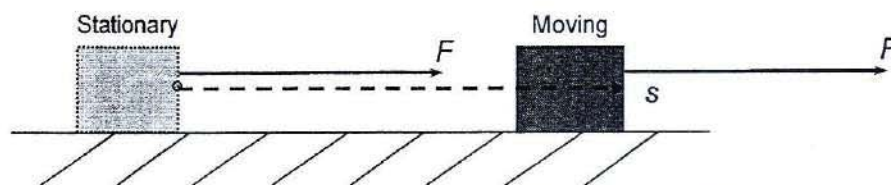
5.1 Work

The meaning of 'work' in physics differs from everyday usage. In Physics, it has a specific meaning that is clearly defined. It describes an effect that occurs when a force interacts with a body that undergoes displacement. It is about the transfer and transformation of energy between one body and another.

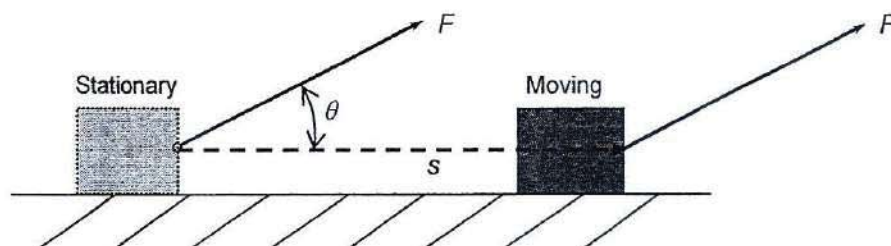
A person holding a heavy book *at rest* may say that he is doing hard work (in the physiological and psychological sense) but he is doing no work at all in the eyes of a physicist. There is no contradiction here. The physicists refer to the work done by a force that is exerted by the person on the book. The physicist is concerned strictly with 'the cause' (a particular force) and its effect (work done) on a particular body.

5.1.1 Work done by a constant force

When a force acts through a distance, we say "the force does work". In the simplest case, consider a constant force F acting on a body which undergoes displacement s in the direction of the force.



The work done BY the force ON the body is defined as the **product of the magnitude of the force F and the displacement s in the direction of the force.**



For the general case when the displacement is not in the direction of the constant force, we can also determine the work done by the force as the product of the displacement and the component of the force in the direction of the displacement. As can be seen below, both lead to the same mathematical expression.

$$W = F s_{\parallel} = F (s \cos \theta)$$

Alternatively

$$W = F_{\parallel} s = (F \cos \theta) s$$

where

F is the magnitude of constant force

s is the displacement

θ is the angle between the force and displacement

Kiasu
ExamPaper

Notes:

- Work is a **scalar** quantity. The unit of work is the **joule (J)**.
- The force F acts on a *particular point* on the object. The displacement in this expression refers to the displacement of *this point* alone. This point is sometimes called the point of application of the force.
- It is assumed that the object is *not rotating*, or *changing shape* (i.e. it is rigid), so all parts of the object move with the same displacement. The physics of rotating or non-rigid bodies are beyond the A-level syllabus.
- Where there are more than one force acting on a body, work done by the individual forces can be calculated separately.

Concept Check 1

- (a) When a man is standing and holding a heavy load, is he doing mechanical work on the load? Explain.
 (b) As the man is walking around at a constant speed and holding up the load, is any mechanical work being done on the load? Explain why.



a) No. There is no displacement.

b) No. F & s are $\perp$.

$$W (by F) = F s \cos 90^\circ = 0 \text{ J.}$$

Concept Check 2

Under what circumstance is the work done by a force negative?

$$W = F s \cos \theta$$

when $270^\circ > \theta > 90^\circ$, $\cos \theta$ is -ve.

$\therefore$ Dnp in energy

when $\theta = 180^\circ$, $\cos \theta = -1$

Example 1

A lawn mower is pushed by a force P of 20.0 N at 30° to the horizontal, against a horizontal resistive force D of 8.0 N. It moves a horizontal distance of 10.0 m. Calculate

- the work done by the force P on the lawn mower.
- the work done by the resistive force D .
- the work done by the normal contact force from the ground and the weight of the lawn mower.
- the change in kinetic energy of the mower.

(i). $W_P = P s \cos \theta = 20.0 \times 10.0 \times \cos 30^\circ = 173 \text{ J}$

(ii). $W_D = D s \cos \theta = 8.0 \times 10.0 \times \cos 180^\circ = -80 \text{ J}$

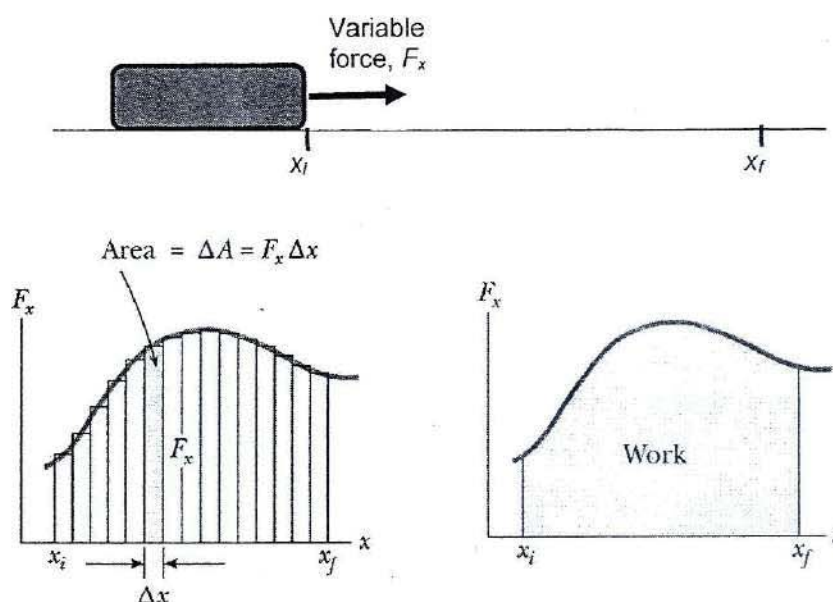
(iii). $N \perp s$ and $W \perp s$, so 0 work done by N & W .

(iv). Net work done on the lawn mower is

$$\Delta KE = W_P + W_D = 173 + (-80) = 93 \text{ J}$$

5.1.2 Work done by a variable force

Suppose an object is being displaced along the x -axis under the action of a force F_x that acts in the x direction and varies with position. The object is displaced in the direction of increasing x from $x = x_i$ to $x = x_f$. In order to calculate work done by the **variable force** F_x , we first need to know how the force applied F_x varied with displacement. Let's imagine we were able to use a force sensor and a data logger to measure this variation and that we were able to get the plot shown below.



The total displacement x can be regarded as the sum of many very small steps Δx . During each small step, the force is practically constant.

The total work done = sum of work done for each step

$$W = \Delta W_1 + \Delta W_2 + \Delta W_3 + \dots$$

$$= F_1 \Delta x + F_2 \Delta x + F_3 \Delta x + \dots$$

As the width of the small steps, Δx , gets really small, the sum of area of these strips will just be the area under the graph of F_x against x .

$$W = \int_{x_i}^{x_f} F_x \, dx$$

In other words, the work done by a variable force acting on an object that undergoes a displacement is equal to the area under the force-displacement graph.

5.2 Energy

In principle, you can solve problems regarding the motion of an object by applying Newton's three laws of motion using vectorial methods. But some problems aren't as simple as they look. How would you determine the speed of a pinball launched by a compressed spring that applies a varying force? How would you determine the motion of one particular molecule in a volume of gas with billions of molecules in random motion? The simple vectorial method that you have learned earlier does not present easy solution where variable forces are involved or where 3 or more bodies are interacting in a system.

The concept of energy provides an **alternative method** for dealing with problems in mechanics. It unifies every area of physics even though it is difficult to define just what energy is. Although it is often convenient to classify energy as being chemical or electrical or thermal etc., there are fundamentally only two types of energy – kinetic energy and potential energy from the view point of mechanic.

A body, which is capable of doing work, is said to possess energy. The amount of energy that a body has is equal to the amount of work that it can do.

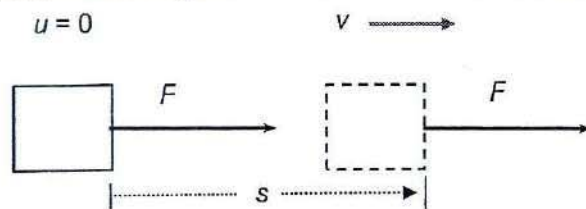
Hence, the energy of a system is a measure of its capacity of to do work.

5.3 Mechanical Energy

The total **mechanical energy** of a system is the sum of all types of **kinetic energy** and **potential energy** of the system.

5.3.1 Kinetic energy (Derivation of $E_k = \frac{1}{2}mv^2$)

Kinetic energy of a body is a measure of the energy possessed by the body by virtue of its motion. Suppose an object is being displaced along the x-axis under the action of a constant force,



we can infer the formula for kinetic energy from the amount of work that is done by an external force to bring a body from rest to its state of motion.

$$E_k = F s$$

By Newton's second law:

$$E_k = (ma)s$$

Using the equations of motion for uniform acceleration:

$$E_k = ma \left(ut + \frac{1}{2}at^2 \right)$$

$$= ma \left(\frac{1}{2}at^2 \right)$$

$$E_k = \frac{1}{2}ma^2t^2$$

$$= \frac{1}{2}m \left(\frac{v-u}{t} \right)^2 t^2$$

$$= \frac{1}{2}m \left(\frac{v-0}{t} \right)^2 t^2$$

$$= \frac{1}{2}mv^2$$



Example 2

In the earlier derivation of $E_k = \frac{1}{2}mv^2$, we conveniently assumed the body was initially at rest. Can you show that in general, work done by a constant force is the gain in kinetic energy of the body?

$$W = Fs = \frac{1}{2}m(v^2 - u^2) = \frac{1}{2}mv^2 - \frac{1}{2}mu^2$$

∴ work done = gain in K.E

5.3.2 Potential energy

An object can have energy by virtue of its position or the arrangement of the system that it is part of. The potential energy of a system represents a form of stored energy, which can be fully recovered and converted into kinetic energy.

Potential energy of a body can be defined as the amount of work that was done on it to give it that position.

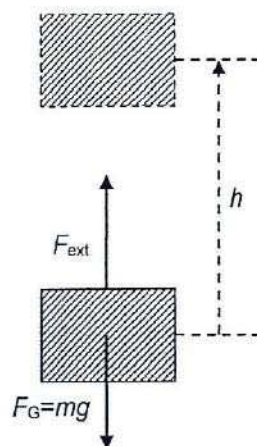
There are three kinds of potential energy. They are summarized in the table below.

Elastic PE	Gravitational PE	Electrical PE
Interaction between charges on atoms and molecules	Interaction between masses	Interaction between charges
Involves attractive and repulsive forces	Involves only attractive forces	Involves attractive and repulsive forces

5.4 Gravitational Potential Energy

To move objects around near the surface of the earth we need to use forces to overcome the Earth's attraction. The work done by exerting forces to overcome this attraction and place objects in positions further and further from the surface of the earth manifests itself as an increasing gravitational potential energy associated with that object.

The **gravitational potential energy** (G.P.E) of a body is a measure of the energy it possesses by virtue of its position in a gravitational field.



We can infer the formula for G.P.E. from the amount of work that is done **by an external force** F_{ext} to lift a body through a vertical distance of h at constant speed.

$$W = F_{\text{ext}} s$$

To just overcome the gravitational pull, $F_{\text{ext}} = mg$.

$$\begin{aligned} W &= (mg)(h) \\ &= mgh \end{aligned}$$

Hence,

$$\text{gain in G.P.E} = mgh$$

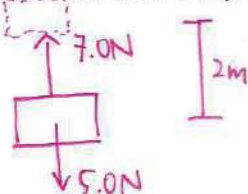
Note:

This derivation applies only to objects that stay near to the surface of the Earth during the motion, and h is small relative to the diameter of the Earth so that the gravitational pull on the body (i.e. its weight mg) is practically constant during the motion.

Example 3

A mass of weight 5.0 N is lifted by a vertical force of 7.0 N through a vertical distance of 2.0 m.

- Calculate the work done by the lifting force.
- Calculate the change in Gravitational Potential Energy, assuming it was initially at rest.
- Calculate the final speed of the mass.



$$\begin{aligned} \text{a) work done} &= [7.0\text{N} - (5.0\text{N})] \times 2\text{m} \\ &= 24\text{J} \quad 7.0\text{N} \times 2\text{m} = 14\text{J} \end{aligned}$$

$$\begin{aligned} \text{b) } \Delta \text{G.P.E} &= mgh \\ &= 5 \times 2 \\ &= 10\text{J} \end{aligned}$$

KIASU
ExamPaper

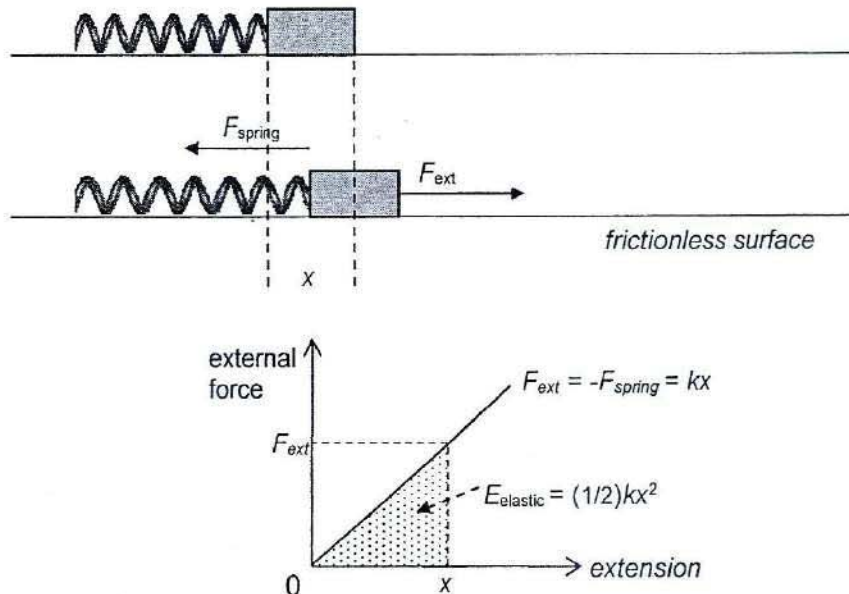
c)

$$\begin{aligned} \text{Work done} &= \Delta \text{KE} + \Delta \text{GPE} \\ 14 &= \frac{1}{2} \times \frac{5.0}{9.81} \times v^2 + 10 \\ v &= 4.0\text{m/s} \end{aligned}$$

5.5 Elastic Potential Energy

Forces need to be exerted to alter the shape (length) of a spring. Hence for each particular extension of the spring, there is an associated elastic potential energy value. For any particular displacement, this depends on the stiffness of the spring. A stiffer spring requires larger forces to deform it so that you will end up having done more work to attain a particular extension and hence will have a larger elastic potential energy associated with that extension.

The external force which is required to change the position of an object near the earth's surface is always equal to the weight of the object and hence constant. **Changing the length of a spring, however, requires an external force that increases in proportion to the extension.**



Let a light spring be stretched through an extension x , very slowly (at practically zero velocity) by applying an external force F_{ext} .

If the spring obeys Hooke's law, the spring force F_{spring} is proportional to its extension x , i.e. $F = -kx$, where k is the spring constant. To just overcome the spring force, the external force $F_{ext} = -F_{spring}$.

$$\begin{aligned}
 W &= \int_0^x F_{ext} dx = \text{Area under the force-extension graph} \\
 &= \frac{1}{2} F x \\
 &= \frac{1}{2} (kx) (x)
 \end{aligned}$$

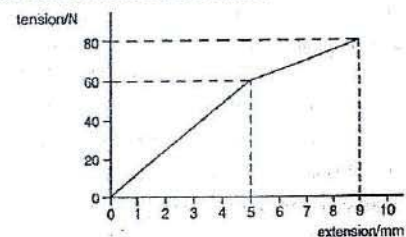
$$E_{elastic} = \frac{1}{2} kx^2$$

Hence, elastic potential energy $E_{elastic} = \frac{1}{2} Fx = \frac{1}{2} kx^2$

Example 4 (J99/II/23)

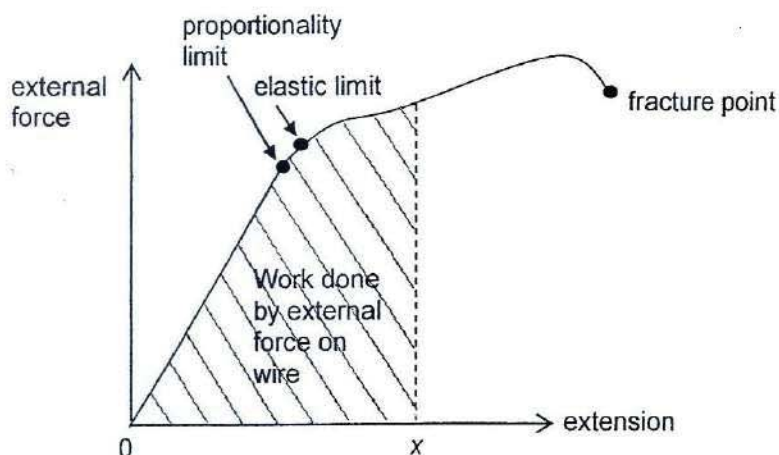
A sample is placed in a tensile testing machine. It is extended by known amounts and the tension is measured. What is the work done on the sample when it is given a total extension of 9.0 mm?

KIASU
ExamPaper
0.43J



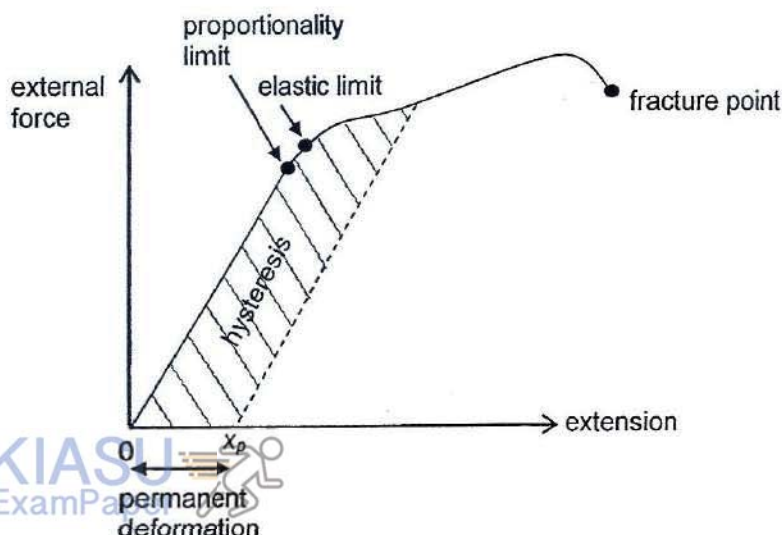
5.5.1 Work done by External Force on a Deformed Material

An elastic material (e.g. a wire) usually obey Hooke's Law closely unless the deformation is excessively large. If the material is stretched beyond its proportionality limit, it will no longer obey Hooke's Law and the force-extension graph will not be a straight line from that point onwards. If an elastic material (e.g. a wire) is stretched further **beyond its elastic limit** (also known as yield point), it will **suffer a permanent deformation** and not snap back to its original length when the external force is removed. If the wire is stretched even further, it will eventually hit fracture point and the wire will simply snap. The force-extension graph will look like the one shown in the figure below.



In such cases, to determine work done by external force on wire, one cannot simply use the equation $E = \frac{1}{2} kx^2$ since Hooke's Law is no longer obeyed. Instead, the work done can simply be worked out by calculating the area under the force-extension graph (shaded area).

When the wire is **stretched beyond its elastic limit** and the external force is subsequently removed, the extension is only partially recovered. Work done by external force on wire will be more than the elastic potential energy. Energy is therefore lost as heat and this phenomenon is known as hysteresis.



Try this out ! The effect of hysteresis is usually very small for metals, but is noticeable for polythene, glass and rubber. You can easily investigate this using a rubber band. By simply stretching it and then holding it against your lips you can detect a rise in temperature!

5.6 Principle of Conservation of Energy

The importance of the energy idea stems from the principle of conservation of energy:

Energy can be converted from one form to another but cannot be created or destroyed.

Example 5

A motorcyclist leaps across a canyon by driving horizontally off a cliff at a speed of 38.0 m s^{-1} . Ignoring air resistance, find the speed with which the cycle strikes the ground on the other side.



$$\begin{aligned}\frac{1}{2}mv^2 &= 35 \times m \times 9.81 \\ v^2 &= 70 \times 9.81 \\ v &= 46 \text{ m/s}\end{aligned}$$

Example 6

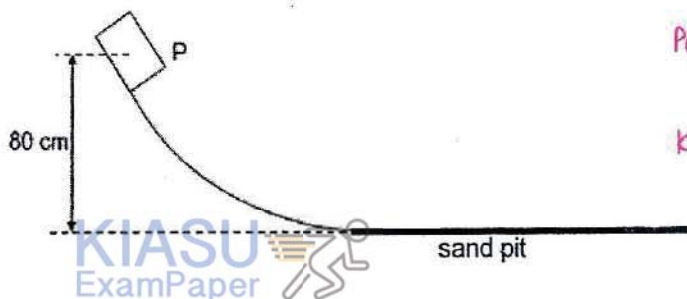
A car of mass 800 kg moving at 30 km h^{-1} along a horizontal road is brought to rest by a constant retarding force of 5000 N . Calculate the distance the car moves before coming to rest.

8.33 m/s

$$\begin{aligned}\text{K.E before} + \text{work done by retarding force} &= \text{KE after} \\ \frac{1}{2}mu^2 + (-F)s &= 0 \\ s &= 5.6 \text{ m}\end{aligned}$$

Example 7

A block of mass 4.0 kg is released at point P on a curved frictionless track as shown. The block slides down a vertical distance of 80 cm and enters a sand pit. Determine the distance the block can move along the sand pit before coming to rest if the friction acting on the block is 85 N .



Potential energy
↓
Kinetic energy
↓
Internal energy

$$\begin{aligned}\text{PE}_{\text{at P}} + \text{work done on the block by friction} &= 0 \\ \text{PE}_{\text{at P}} &= 31.392 \text{ J} \\ W &= -F_s \\ s &= 37 \text{ cm}\end{aligned}$$

5.7 Power

The same amount of work is done in raising a given body through a given height, regardless whether it takes one second or one year to do so. However, the rate at which work is done is greater for the former than that for the latter.

Power is defined as the rate at which work is done.

$$\text{Instantaneous power, } P = \frac{dW}{dt}$$

$$\text{Average power, } \langle P \rangle = \frac{\Delta W}{\Delta t}$$

Notes:

- The SI unit of power is the **watt** (W). Power is a **scalar** quantity.
- Work can also be expressed in units of (power x time). This is the origin of the term *kilowatt-hour*.
One *kilo-watt hour* is the work done in 1 hour by an agent working at a constant rate of 1 kW.

Example 8

A 1.10×10^3 kg car, starting from rest, accelerates for 5.00 s. The magnitude of the acceleration is $a = 4.60$ m s⁻². Determine the average power generated by the net force that accelerates the vehicle.

$$W = KE_{\text{final}} - KE_{\text{initial}} = KE_{\text{final}}$$

$$KE_{\text{final}} = \frac{1}{2}mv^2 = \frac{1}{2}m(at)^2 = 290950 \text{ J}$$

$$P = \frac{290950}{5} = 5.82 \times 10^4 \text{ W}$$

5.7.1 Relationship between Power, Force and Velocity

Consider a force F that acts on a body for a small time interval Δt . The body moves a small displacement Δx in the direction of the force.

Work done by the force F during Δt , $\Delta W = F \Delta x$

Power delivered by that force F during the time interval Δt

$$\begin{aligned} P &= \Delta W / \Delta t \\ &= (F \Delta x) / \Delta t \\ &= F(\Delta x / \Delta t) \\ &= Fv \end{aligned}$$

where v is the instantaneous velocity of the body.

When a force F acts on a body that is moving with velocity v , in the direction of the force, it delivers power to the body at the rate given by $P = Fv$.

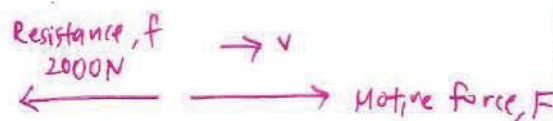
Notes:

- A force F that acts in the direction of the displacement Δx , does positive work on the body. It transfers energy to the body at a rate proportional to the speed of the body.
- If F or v varies with time, then P may not be constant. However, the product Fv at any moment will give the instantaneous power of F at that instant.
- Where F acts against the motion of the body, it does negative work on the body. It extracts energy out of the body at a rate equal to the magnitude of Fv .

Example 9

A car is moving with a constant velocity of 20.0 m s^{-1} along a straight horizontal road. The total resistive forces acting on the car is 2000 N . What is the power developed by the engine of the car?

$$P = Fv = 4.00 \times 10^4 \text{ W}$$

**5.8.2 Efficiency (of energy conversion), η**

No practical engine can transform energy from one form to another without some energy going to internal energy and other non-useful forms of energy.

The efficiency of any energy conversion process is the ratio of

- the useful energy output to the total energy input, or
- the useful power output to the power input

$$\text{efficiency} = \frac{\text{useful energy output}}{\text{total energy input}}, \quad \eta = \frac{E_{\text{output}}}{E_{\text{input}}} = \frac{P_{\text{output}}}{P_{\text{input}}}$$

Note:

η (Greek small letter, Eta) is a ratio between two quantities of the same type. It has no unit.

Example 10

A small motor is used to raise a weight of 2.0 N at constant speed through a vertical height of 80 cm in 4.0 s . The efficiency of the motor is 20% . What is the electrical power supplied to the motor?

Useful energy output $E_{\text{output}} = \text{gain in GPE}$

$$E_{\text{output}} = mgh = 2.0 \times 0.80 = 1.6 \text{ J}$$

Useful power output $P_{\text{output}} = E_{\text{output}} / t$

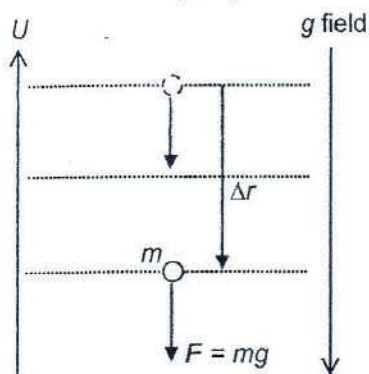
$$P_{\text{output}} = 0.40 \text{ W}$$

$$\eta = P_{\text{output}} / P_{\text{input}}$$

$$P_{\text{input}} = 2.0 \text{ W}$$

Appendix 1:**Relationship between field force and potential energy (to be covered in later chapters)**

We have seen that in general, potential energy is a function of position such that the difference between its value at an initial and final position is always equal to the work done in moving an object between these positions. We will use this concept to arrive at a general rule for the relationship between a potential energy function (how the potential energy varies with displacement) and the size of the force generated by the field. We shall for simplicity's sake consider a uniform gravitational field.



Consider a mass that falls freely under the pull of gravity F , through a vertical distance Δr in the direction of F .

Work done by the gravitational force, $W = F \Delta r$

Work done = Difference in Potential Energy = $-\Delta U$ (Decrease in P.E.)

$$\therefore F \Delta r = -\Delta U \quad \Rightarrow F = -\frac{\Delta U}{\Delta r}$$

This concept can be generalised for any conservative field in which energy is not lost. A force is conservative if the work it does on an object moving between two points is independent of the path taken between the points. Examples of conservative forces are gravitational force, electrostatic force and force of a spring.

Conservative fields can have a potential energy function associated with it. Work done in bringing an object from one point to another is simply the difference in the potential energy between those two points.

In any **conservative and uniform field**, the force exerted by the field is related to the potential energy U of the system by

$$F = -\frac{\Delta U}{\Delta r}$$

The **magnitude of the field force** at a point in the field is given by the **potential energy gradient** at that point.

The **direction of the field force** is in the **direction of decreasing potential energy**.

In general, the concept can be generalised to include non-uniform fields:

$$F = -\frac{dU}{dr}$$

Topic 6: Circular Motion

H2 Physics C1 2016

Chapter 6: MOTION IN A CIRCLE

H2 Physics Syllabus 9749

Content

- Kinematics of uniform circular motion
- Centripetal acceleration
- Centripetal force

Learning Outcomes

Candidates should be able to:

- Express angular displacement in radians.
- Show an understanding of and use the concept of angular velocity to solve problems.
- Recall and use $v = r\omega$ to solve problems.
- Describe qualitatively motion in a curved path due to a perpendicular force, and understand the centripetal acceleration in the case of uniform motion in a circle.
- Recall and use centripetal acceleration $a = r\omega^2$, $a = v^2/r$ to solve problems.
- Recall and use centripetal force $F = mr\omega^2$, $F = mv^2/r$ to solve problems.

Further Readings / References

- 1) University Physics with Modern Physics, 11th Ed., *Young & Freedman*, Chapter 5, Pg 181.
- 2) Advanced Level Physics, 6th Ed., *Nelson & Parker*, Chapter 2, Pg 48.
- 3) College Physics, 8th Ed., *Young & Geller*, Chapter 6, Pg 161.
- 4) Physics for Scientists and Engineers with Modern Physics, 6th Ed., *Serway Jewett*, Chapter 6, Pg 150.

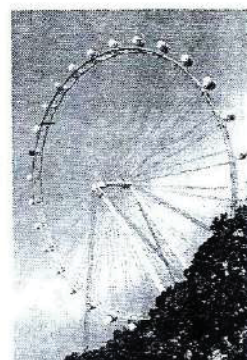
6.1 Introduction

Circular motion is a type of motion we often come across in our daily lives, e.g. Cars navigating a bend, the capsules of a Ferris wheel, children on a merry-go-round, clothes in a spin dryer.

On a larger scale, in astronomy, we know that the moon circles the Earth, which circles the Sun, which circles the centre of the Milky Way.

Have you ever wondered?

- Why must a cyclist lean into a turn?
- Why must an airplane tilt when executing a turn?
- Why people in a roller coaster don't fall out at the top of the loop? (No, it is not because of the belts)



The exploration of the fascinating world of circular motion begins here.

A Short Revision on basic ideas of Newton's Laws of Motion

Newton's 1st law:

Describes the motion of any object not subjected to a net external force, specifically that

- an object at rest remains at rest, or
- an object continues with constant speed in a straight line.

In the presence of a net external force, a constant mass experiences an acceleration governed by *Newton's 2nd Law*:

$$\vec{a} = \frac{\vec{F}_{net}}{m}$$

Note:

- Forces which are parallel to the motion of an object result in its motion either speeding up or slowing down, with the direction of motion remaining along the original axis.
- Forces which are perpendicular to the motion of the object cannot affect its speed, but instead will change its direction of motion.
- Forces which are neither wholly parallel nor perpendicular to the direction of motion will affect both the speed and direction of the object upon which the force acts. (Consider projectile motion).

In this chapter you will learn that objects will move in circular paths when they experience a force or forces that always act perpendicularly to their motion.

6.2 Kinematics of Circular Motion

When an extended object such as the wheel of a bicycle is rotating, different parts of the wheel will travel different distances in the same time. The outer parts of the wheel will travel larger distances and hence will have higher speeds than the inner parts of the wheel. However all parts of the wheel will pass through the same angle in a given time. Hence the speed of rotation of a wheel is better expressed as an angular speed as this is the same for all parts of the wheel.

6.2.1 How angle is measured

The S.I. unit for angles is not the degree but the radian. Angle measured in radian is actually the ratio of two lengths: arc length and radius. The *arc length* s is the distance travelled along the circular path, and the angle θ is said to *subtend* the arc length at the circle centre (Fig. 1).

Hence, note that θ is actually a dimensionless quantity since it is a ratio of 2 lengths.

$$\theta \text{ (in rad)} = \frac{\text{arc length}}{\text{radius}} = \frac{s}{r}$$

$$s = r\theta$$

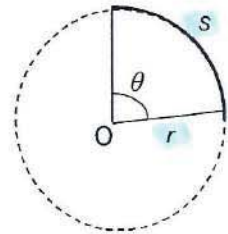


Figure 1

One radian is the angle subtended at the centre of the circle by an arc equal in length to the radius of the circle.

If circumference of a circle $s = 2\pi r$, $\theta = s/r = 2\pi \text{ rad}$.
 Since for a complete circle, the angle subtended is 360° ,

$$360^\circ \equiv 2\pi \text{ rad}$$

We can deduce that $1 \text{ radian} = \left(\frac{360}{2\pi}\right)^\circ = 57.3^\circ$

Conversion between degrees and radians:

$$180^\circ \equiv \pi \text{ rad}$$

$$Z \text{ (degrees)} \equiv Z \times \frac{\pi}{180} \text{ (radians)}$$

6.2.2 Angular Displacement ($\Delta\theta$)

Angular displacement $\Delta\theta$ is the angle an object turns about a fixed point. In circular motion, the fixed point is taken to be the centre of the circle (Fig. 2).



$$\Delta\theta = \theta - \theta_0$$

If $\theta_0 = 0$, then $\Delta\theta = \theta$.

S.I. unit of $\Delta\theta$: radian (rad)

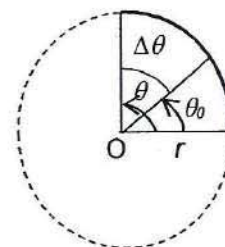


Figure 2

6.2.3 Angular Velocity (ω)

The description of circular motion in angular form is analogous to the description of linear motion. Angular velocity ω (pronounced as *omega*) is defined as the rate of change of angular displacement.

$$\text{Average angular velocity, } \bar{\omega} = \frac{\text{Angular displacement}}{\text{Elapsed time}}$$

$$\bar{\omega} = \frac{\theta - \theta_0}{t - t_0} = \frac{\Delta\theta}{\Delta t}$$

The *instantaneous angular velocity* ω is the angular velocity that exists at any given instant. Analogous to instantaneous linear velocity,

$$\text{Instantaneous angular velocity, } \omega = \frac{d\theta}{dt}$$

If an object has a **constant** angular velocity, the instantaneous value and the average value are **the same** (i.e. $\omega = \bar{\omega}$).

Note:

- Magnitude of the instantaneous angular velocity is called the *instantaneous angular speed*.
- S.I. unit of ω : rad s^{-1}

6.2.4 Relationship between Tangential Speed v and Angular Speed ω

A particle moving in a circle has an instantaneous velocity tangential to its circular path. For a **constant** angular speed, the particle's **orbital or tangential speed** v is also constant. (v is also known as *linear speed*.)

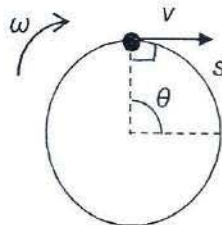


Figure 3

Since $s = r\theta = r(\omega t)$. Thus $s = r\omega t$

The arc length, or distance, s is also given by $s = vt$

Thus,

$$\text{tangential speed, } v = r\omega$$

Alternatively,



Taking derivative with respect to time,

$$\frac{ds}{dt} = r \frac{d\theta}{dt} \rightarrow v = r\omega$$

Unit of linear/orbital speed: m s^{-1}

6.2.5 Period (T) and Frequency (f)

Period (T) is the time it takes for an object in circular motion to make one complete revolution, or cycle.

Frequency (f) is the number of revolutions, or cycles, made per unit time. The unit of frequency is s^{-1} , which is called the *hertz* (Hz) in the SI.

$$\text{frequency, } f = \frac{1}{T}$$

Since an angular displacement of 2π rad is travelled in 1 period,

$$\text{angular velocity, } \omega = \frac{2\pi}{T} = 2\pi f$$

Example 1

$$r = \frac{120\text{mm}}{2} = 60\text{mm} = 0.060\text{m}$$

A Compact Disc player is spinning a standard CD (120 mm diameter) at 300 revolutions per minute (RPM). Determine

- the frequency of the rotation (in revolutions per second)
- the period of the rotation
- the angular speed

$$f = \frac{300}{1\text{min}}$$

$$(i). f = \frac{300\text{ rev}}{1\text{ min}} = \frac{300\text{ rev}}{60\text{ sec}} = 5.00\text{ Hz}$$

$$(ii). T = \frac{1}{f} = \frac{1}{5.00} = 0.200\text{ s}$$

$$(iii). \omega = 2\pi f = 2\pi (5.00) = 31.4\text{ rad/s}$$

Example 2

Dishes are placed on a Lazy Susan at a reunion dinner during Chinese New Year. When the Lazy Susan is spun around to serve the guests, what can be said about the dishes in the inner circle compared to those in the outer circle in terms of their speed and angular velocity?



slower speed, but the same angular velocity

$$v = r\omega$$

↓ same angular velocity

6.3 Uniform Circular Motion

Uniform circular motion is the motion of an object travelling at a **constant** (uniform) speed in a circular path.

An object moving in a uniform circular motion (Fig. 4) has a *constant angular velocity* (ω). The **magnitude** of the velocity vector (or the linear speed, v) is **constant** but the **direction** of the velocity vector is **changing**. Thus, the *linear velocity vector is continuously changing* (v_1 , v_2 & v_3) as the object moves around the circle. Since there is a change in the velocity vector with respect to time, the object should be under acceleration.

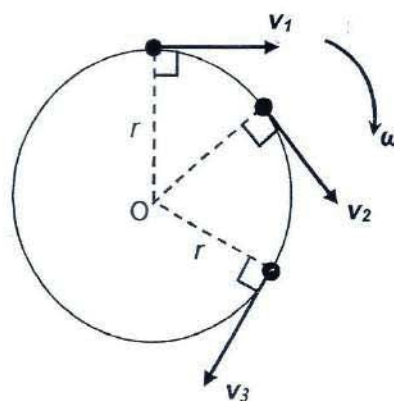


Figure 4

Derivation of the equation of centripetal acceleration:

In this derivation we only consider the kinematics of uniform circular motion to arrive at a general expression for the acceleration an object is experiencing when it moves in a circle at constant speed.

Consider a particle moving in a circle of radius r with uniform angular speed (Fig. 5a). At one instant the particle is at A, and its instantaneous velocity is v in the direction AA' .

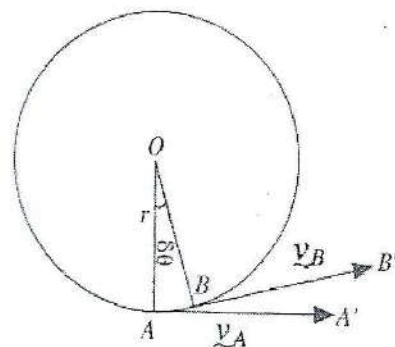


Figure 5a

A very short time later δt later, the particle has moved to B, a distance $r\delta\theta$ along the arc, where $\delta\theta$ is a small angle and its velocity is v in the direction BB' at this time.

With reference to the vector triangle in Fig. 5b, the arc length can be approximated to a chord.

For small angle $\delta\theta$,

$$\delta v \approx v\delta\theta$$

$$\delta v/\delta t \approx v\delta\theta/\delta t$$

$$\frac{dv}{dt} = v \frac{d\theta}{dt}$$

$$a_c = v\omega$$

Similarly,

$$a_c = (r\omega)\omega = r\omega^2$$

$$a_c = v\left(\frac{v}{r}\right) = \frac{v^2}{r}$$

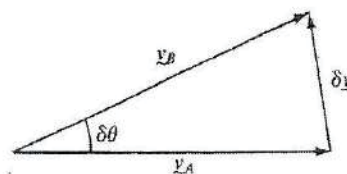


Figure 5b

As δt (or $\delta\theta$) (Fig. 5b) becomes smaller, the angle between δv and either v_A or v_B will tend to 90° . Since either v_A or v_B are the tangents to the circular path, this means δv will point towards the centre of the circular path. Similarly, its acceleration will point towards the centre of the circular path because δv and its acceleration are in the same direction. This type of acceleration in uniform circular motion is called **centripetal acceleration**,

The centripetal acceleration is directed **radially inward** towards the centre of the circular path. The direction of the centripetal acceleration is continuously changing.

For an object in **uniform circular motion**, there is **no** acceleration component in the tangential direction (i.e. angular acceleration), or else the magnitude of the velocity vector (tangential or linear speed) would change.

Hence,

$$\text{Centripetal acceleration, } a_c = \frac{v^2}{r} = r\omega^2 = v\omega$$

Conceptual Questions:

- (a) Is an object travelling at non-uniform speed accelerating?

Yes, acceleration involves any change in velocity (magnitude and/or direction), and the speed is changing in this case.

- (b) Can an object travelling at constant speed be accelerating?

Yes, although its speed may be constant, its direction of motion, thus its velocity, may change, hence there must be an acceleration.

Centripetal Force (F_c)

In the last section we discovered both the magnitude and direction of the acceleration associated with circular motion. We now move into the dynamics of circular motion, quantifying and identifying forces that can constrain objects to adhere to such circular paths.

To provide acceleration, there must be a net force. Thus to produce a centripetal (inward) acceleration, there must be a resultant force towards the centre of the circular motion. As the resultant force is centre-seeking, we call it the centripetal force.

From Newton's second law ($\Sigma F = ma$),

$$F_c = ma_c = \frac{mv^2}{r} = mr\omega^2 = mv\omega, \text{ and it is always directed toward the centre of the circle.}$$

Note: In general, when a force is continuously applied at an angle of 90° to the direction of motion (as is centripetal force), **only the direction** of the velocity changes.

Conceptual Question: Is there work done by a centripetal force?

A centripetal force does no work because there is no displacement in the direction of the centripetal force. The centripetal force is always perpendicular to the direction of motion (displacement).

Note: For uniform circular motion,

- Centripetal force **does not** denote a new and separate force created by nature.
- Centripetal force is provided by the **net force** pointing toward the centre of the circular path, and this net force is the **vector sum of all the force components that point along the radial direction**.

Centripetal force is not a new type of force. "Centripetal" is just an adjective employed to describe the action of a force that is causing circular motion. E.g. the force of friction acting between the tires of a car and the road can provide a force which will allow the car to be steered around a bend. In this case it would be legitimate to call the friction a centripetal force because it acts in such a way as to cause circular motion. Friction is also the force that exists between the brake disc attached to the wheels of a car and the brake pads attached to the car itself which slow the car down when required. The friction is now acting in such a way as to cause linear deceleration of the car and could now be better described as a braking or resistive force but definitely not a centripetal force. Moreover the term "centripetal force" should never appear in a Free Body Diagram of an object in circular motion.

6.4 Problem solving for Circular Motion

6.4.1 General strategy for solving circular motion problems

1. Identify the object undergoing circular motion (and the relevant known quantities such as m , v , r , ω ...).
2. Draw a **free body diagram** of the object, carefully identifying the individual forces *experienced by* the object.
3. **Resolve forces** into relevant perpendicular "directions" or use triangle of forces to arrive at a net force acting on the object.
4. Equate this net force to $\frac{mv^2}{r}$ to ascertain any unknown.



6.4.2 Horizontal Circular Motion

Example 3 - Conical pendulum

In a conical pendulum system, a small pendulum bob of mass 0.50 kg is rotating in a fixed horizontal plane. The string is 30 cm long and makes an angle of 15° to the vertical.

Calculate the (a) tension in the string,
(b) linear speed of the bob, and
(c) period of oscillation of the bob.



Free-body diagram and vector triangle:

$$\begin{aligned} \text{Vertical: } T \cos \theta &= mg \\ \text{Horizontal: } T \sin \theta &= \frac{mv^2}{r} \end{aligned}$$

Calculations:

a) $T \cos 15^\circ = mg$
 $= (0.50)(9.81)$
 $T = 5.08 \text{ N}$

b) $\tan \theta = \frac{v^2}{rg}$
 $\tan(15^\circ) = \frac{v^2}{0.30 \sin(15^\circ) \cdot 9.81}$
 $\Rightarrow v = 0.45 \text{ m/s}$

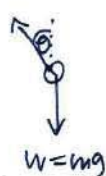
c) $v = r\omega$
 $= r \frac{2\pi}{\text{period}}$
 $0.452 = 0.30 \sin 15^\circ \times \frac{2\pi}{\text{period}}$
 $\text{period} = 1.15$

Alternative calculation for (b) and (c):

$$T^2 \sin^2 \theta + T^2 \cos^2 \theta = \left(\frac{mv^2}{r}\right)^2 + (mg)^2$$

$$T^2 = \left(\frac{mv^2}{r}\right)^2 + (mg)^2$$
Example 4

Explain with the aid of a diagram, why the mass at the end of a light inelastic string cannot be whirled in uniform circular motion in such a way that the string is horizontal.



If $\theta = 90^\circ$ (horizontal string), there is no vertical component of the force provided by the string to counteract the weight of the bob thus the bob cannot remain in that position with $\theta = 90^\circ$

$$T \cos \theta = mg$$

As $\theta \rightarrow 90^\circ$,

$$\cos \theta \rightarrow 0,$$

thus $T \rightarrow \infty$



Example 5 - Car going around a bend

$$\frac{45000}{3600} = 12.5 \text{ m/s}$$

(a) A bend in the road has a 50 m radius of curvature. A car of mass 600 kg takes the bend at 45 km h⁻¹.

- What is the centripetal acceleration of the car?
- What is the centripetal force experienced by the car and what provides it?
- If the coefficient of friction, μ , between the tires and the road is 0.50, what is the maximum speed at which the car can turn this bend safely? (Recall: frictional force = μN)
- What will happen to the car if the driver decided to take the bend at 60 km h⁻¹ instead?

$$(i). a = \frac{v^2}{r}$$

$$= \frac{12.5^2}{50}$$

$$= 3.125$$

$$= 3.1 \text{ m/s}^2$$

$$(ii). F_c = \frac{mv^2}{r}$$

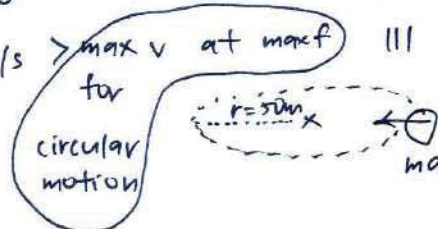
$$= m(3.125)$$

$$= 1875$$

$$7900 \text{ N}$$

$$(iv). 60 \text{ km/h} = \frac{60000}{3600}$$

$$= 16.7 \text{ m/s}$$



∴ The car will skid off the road.

$$(iii) \leftarrow : \Sigma F = ma$$

$$f = \frac{mv^2}{r}$$

$$\mu N = \frac{mv^2}{r}$$

$$\mu mg = \frac{mv^2}{r}$$

$$0.50(9.81) = \frac{v^2}{50} \Rightarrow v = 16 \text{ m/s}$$

- (b) When the car negotiates a corner on horizontal ground, the frictional force between the tyres and the ground is the only force providing the centripetal force. As there is a limit to this frictional force for a particular road surface, there is a maximum speed which the car can make the turn safely, above which skidding will occur.

Hence, some corners (especially at race-tracks) have raised embankments to increase the maximum speed at which a vehicle can take the corner than if on a level road. It does so by making the **normal contact force** contribute a component to the centripetal force.

For an embankment inclined at 20° to the horizontal, find the speed at which the normal contact force is able to **completely provide the centripetal force** (i.e. no frictional force required).

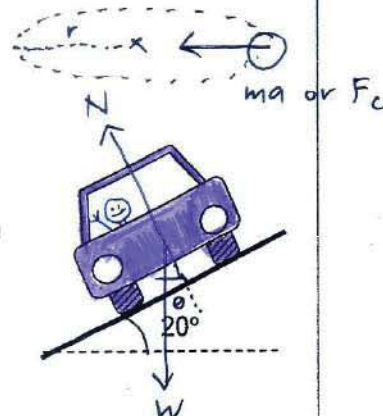
$$\leftarrow : \Sigma F = ma$$

$$N \sin \theta = \frac{mv^2}{r} \quad \text{--- (1)}$$

$$\uparrow : N \cos \theta - W = 0$$

$$N \cos \theta = mg \quad \text{--- (2)}$$

Centre of circular path



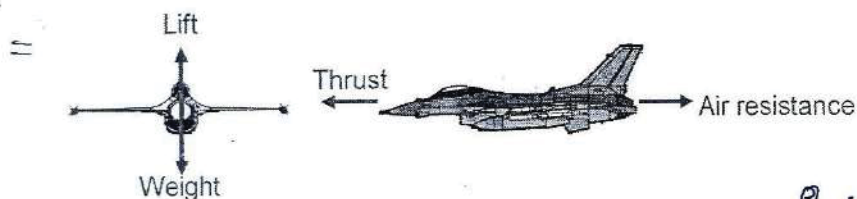
$$\text{Divide (1) by (2)} \quad \frac{N \sin \theta}{N \cos \theta} = \frac{\frac{mv^2}{r}}{mg}$$

$$\tan \theta = \frac{v^2}{rg}$$

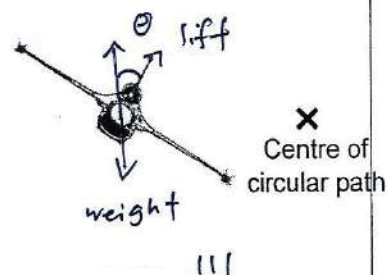
$$\tan 20^\circ = \frac{v^2}{50(9.81)} \Rightarrow v = 13.3 \text{ m/s}$$

Example 6 - Aircraft banking turn

When an airplane flies, it's the flow of air around the wings that generates a force (also known as lift), which is assumed to be perpendicular to the wings. If the plane flies level at constant speed, then the forces acting on it are:

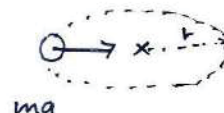


When an airplane executes a turn, it tilts its wings, so that the lift makes an angle θ with the vertical (known as the banking angle), and now has a component in the horizontal direction, that helps the plane to turn. This maneuver is known as **banking**.



Express θ in terms of the linear speed of the aircraft v and the radius of the turn r .

N.B. Lift on the plane is always **perpendicular** to the wings.



$$\rightarrow + : \Sigma F = ma$$

$$L \sin \theta = \frac{mv^2}{r} \quad \text{--- ①}$$

+

$$\uparrow : L \cos \theta - mg = 0$$

$$L \cos \theta = mg \quad \text{--- ②}$$

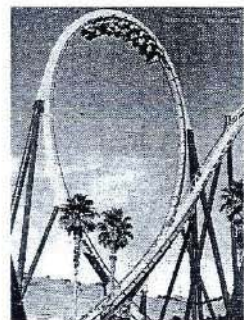
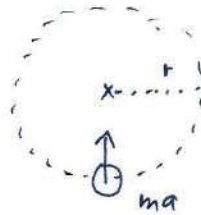
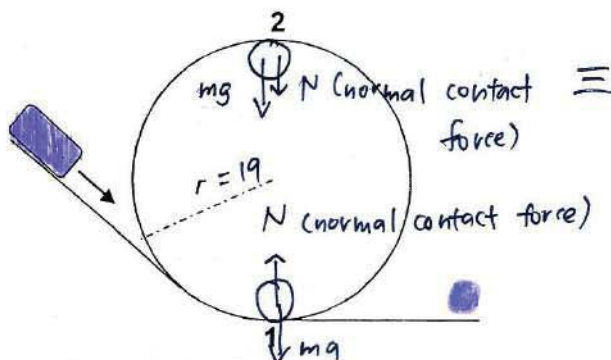
$$\text{①} \div \text{②} \quad \tan \theta = \frac{v^2}{rg}$$

6.4.2 Vertical Circular Motion

When objects travel along a vertical circular path, the speed of the objects is often not constant due to the influence of gravity, resulting in non-uniform circular motion. If, at each point of the motion, we resolve the net force into its tangential and centripetal components, then the former changes only the speed of the motion, the latter changes only the direction of the motion. Calculations-wise, the equation for centripetal acceleration is still $\frac{v^2}{r}$, just that centripetal acceleration must now vary as the speed changes. The **principle of conservation of energy** provides the most convenient method to determine the speed at any position.

Example 7 - Roller coaster

Some roller coasters have several loops along the track. The picture on the right illustrates such a coaster executing a loop-the-loop.

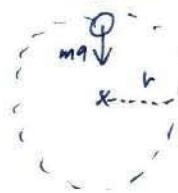


- (a) If a passenger car for a roller coaster enters a loop of radius 19 m at position 1 with a speed of 33 m s^{-1} , determine the normal contact force the track exerts on it at
- the bottom (position 1) and
 - the top (position 2) of the loop.
- (b) Find the minimum speed at which the passenger car must travel at while it is at the top of the loop in order to clear the loop safely.

Assume the total mass of the vehicle and its passengers to be 170 kg and no friction.

$$\begin{aligned} \uparrow : N - mg &= ma \\ &= \frac{mv^2}{r} \\ N &= \frac{mv^2}{r} + mg \end{aligned}$$

$$= 170 \left(\frac{33^2}{19} \right) + 170(9.81) = 1.1 \times 10^4 \text{ N}$$



At the top: Set $N = 0$

$$\begin{aligned} \downarrow : N + mg &= ma \\ + \quad \quad \quad & \\ mg &= \frac{mv^2}{r} \\ 9.81 &= \frac{v^2}{19} \end{aligned}$$

$$v = 13.7 \text{ m/s} = v_2$$

$$\begin{aligned} (1), \downarrow : N + mg &= ma \\ &= \frac{mv^2}{r} \\ N &= \frac{mv^2}{r} - mg \\ &= 170 \left(\frac{18.5^2}{19} \right) - 170(9.81) \end{aligned}$$

KE $\rightarrow$ GPE

Loss in KE = gain in GPE

$$\frac{1}{2}mv_1^2 - \frac{1}{2}mv_2^2 = mgh$$

$$\frac{1}{2}(33)^2 - \frac{1}{2}v_2^2 = 9.81(2 \times 19)$$

$$v_2 = 18.5 \text{ m/s}$$

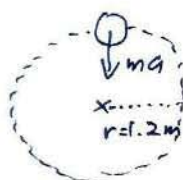
Example 8 – String and bob

A stone of mass 800 g is tied to one end of a string and is whirled in a vertical circle. The string is inextensible and of length 1.2 m. The stone is projected with a certain speed v_A as shown below.

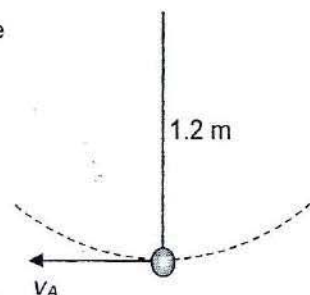
- (i) Determine the minimum speed the stone must have at the top of the circular motion if the string is to be taut at the top.

- (ii) Hence, show that the stone can complete a vertical circular motion if $v_A = 9.0 \text{ m s}^{-1}$.

(i) $w = mg$  $\equiv$



when the string is just taut at the minimum speed,
 $T = 0$



$$\begin{aligned}
 \downarrow : mg + T &= \frac{mv^2}{r} \\
 mg &= \frac{mv^2}{r} \\
 9.81 &= \frac{v^2}{1.2} \\
 v &= 3.43 \text{ m/s}
 \end{aligned}$$

- (ii). Loss in KE = gain in GPE

$$\frac{1}{2}mv_A^2 - \frac{1}{2}mv^2 = mgh$$

$$\frac{1}{2}v_A^2 - \frac{1}{2}(3.43)^2 = 9.81(2 \times 1.2)$$

$$v_A = 7.67 \text{ m/s}$$

Since the stone travels at $v_A = 9.0 \text{ m/s}$ which exceeds the minimum v_A , it will be able to complete the vertical circular motion.

In some other instances of vertical circular motion, the speeds of the objects are made constant (e.g. Ferris wheel). In these cases, the objects are undergoing uniform circular motion. You will encounter some of these applications in the tutorial.

Topic 7: Gravitational Field

H2 Physics C1 2016

Chapter 7: Gravitational Field

H2 Physics Syllabus 9749

Content

- Gravitational field
- Gravitational force between point masses
- Gravitational field of a point mass
- Gravitational field near to the surface of the Earth
- Gravitational potential
- Circular orbits

Learning Outcomes

Candidates should be able to:

- show an understanding of the concept of a gravitational field as an example of field of force and define the gravitational field strength at a point as the gravitational force exerted per unit mass placed at that point.
- recognize the analogy between certain qualitative and quantitative aspects of gravitational and electric fields (will be done in the chapter on electric fields).
- recall and use Newton's law of gravitation in the form $F = \frac{Gm_1m_2}{r^2}$.
- deduce from Newton's law of gravitation and the definition of gravitational field strength, the equation $g = \frac{GM}{r^2}$ for the gravitational field strength of a point mass.
- recall and apply the equation $g = \frac{GM}{r^2}$ for the gravitational field strength of a point mass to new situations or to solve related problems.
- show an appreciation that near the surface of the Earth g is approximately constant and equal to the acceleration of free fall.
- define the gravitational potential at a point as the work done per unit mass in bringing a small test mass from infinity to the point.
- solve problems by using the equation $\phi = -\frac{GM}{r}$ for the gravitational potential in the field of a point mass.
- analyse circular orbits in the inverse square law fields by relating the gravitational force to the centripetal acceleration it causes.
- show an understanding of geostationary orbits and their application.

7.1 Newton's Law of Universal Gravitation

Newton's Law of Universal Gravitation states that:

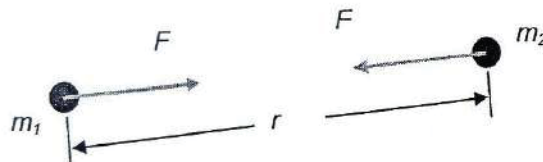
Every particle attracts every other particle with a force that is directly proportional to the product of their masses and inversely proportional to the square of the distance between them.

Consider two particles (point masses) of masses m_1 and m_2 separated by a distance r . Each will exert a force F on the other, given by

$$F = G \frac{m_1 m_2}{r^2}$$

where

F	: gravitational force between the two particles.
m_1, m_2	: masses of the two particles.
r	: distance between the two particles.
G	: gravitational constant.



The two forces form an action-reaction pair and have the following characteristics,

- are equal in magnitude,
- are opposite in direction,
- act on different bodies
- are of the same type (gravitational force).

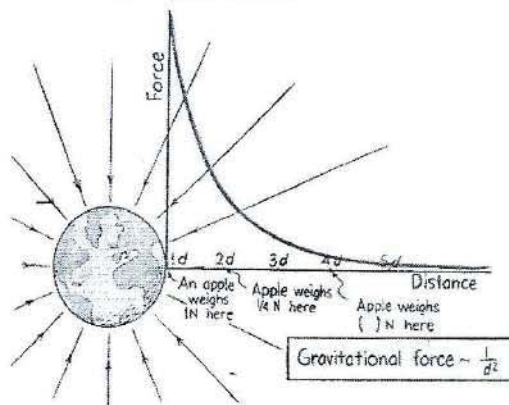
G is called the *gravitational constant* (or constant of universal gravitation), which has been measured experimentally to be: $G = 6.67 \times 10^{-11} \text{ N m}^2 \text{ kg}^{-2}$.

Important points to note about Newton's Law of Gravitation

1. Newton's Law of Gravitation is a **universal** law. It applies everywhere in the universe.
2. **Attractive nature of gravitational force:** Note that the particles in this case are always attracted to each other.
3. The gravitational force is a field force that always exists between two point masses regardless of the medium that separates them. In fact, it would still exist if there were no medium between them.
4. **Inverse-square law:** Newton's Law of Gravitation is an example of an **inverse-square law**, i.e., the force is inversely proportional to the square of the separation of the particles.

$$F \propto \frac{1}{r^2}$$

5. Newton's Law of gravitation force can be applied to spherical bodies as though they were point masses with all their mass concentrated at their geometrical centres.

**Example 1: Does the Earth fall towards the apple?**

Suppose an apple drops from a tree and moves towards the ground. Estimate the acceleration of the Earth towards the apple.

(You may assume that mass of Earth to be $\sim 10^{24}$ kg.)

Solution:

Since both the apple and Earth have mass, they obey Newton's Law of Gravitation.

- The apple is attracted towards the Earth, i.e. Earth exerts a gravitational force on the apple.
- By Newton's 3rd Law, the apple will exert an equal and opposite gravitational force on the Earth:

$$\begin{aligned}
 |F_{\text{Earth}}| &= |F_{\text{apple}}| \\
 &= \text{Weight of the apple} \\
 &= mg = (0.1 \text{ kg})(9.81 \text{ m s}^{-2}) = 0.981 \text{ N} \approx 1 \text{ N}
 \end{aligned}$$

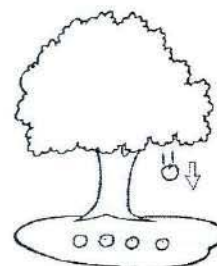
Assuming that no other forces act on the Earth, then the gravitational force acting on the Earth will cause the Earth to accelerate.

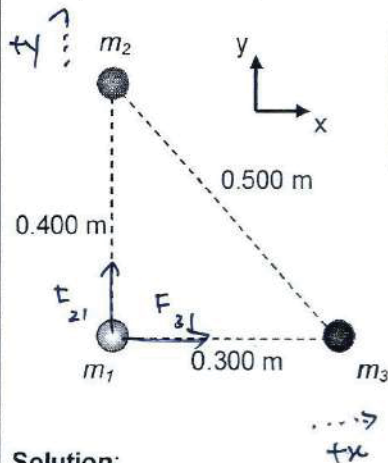
By Newton's 2nd Law,

$$\begin{aligned}
 \text{The acceleration of Earth, } a &= \frac{\text{Resultant force acting on Earth}}{\text{Mass of the Earth}} \\
 &= \frac{F_{\text{earth}}}{m_{\text{earth}}} \\
 &= \frac{(1 \text{ N})}{(10^{24} \text{ kg})} \\
 &\approx 10^{-24} \text{ m s}^{-2}
 \end{aligned}$$

What is the physical significance of this calculation?

Although both the apple and the Earth experience the same magnitude of force, the mass of Earth is much larger. Hence, acceleration a of the Earth is so small that its motion is negligible.



Example 2: Finding Resultant Gravitational Force (Serway and Vuille. p209. Example 7.10)

If several particles are present, each pair will experience a mutual gravitational attraction. The resultant gravitational force on a given particle is the vector sum of the separate attractive forces acting on the particle due to individual particles interacting with it.

Three 0.300 kg billiard balls are placed on a table at the corners of a right triangle.

Prove that the resultant gravitational force on the cue ball m_1 , resulting from the other two balls is 7.65×10^{-11} N, 29.3° anticlockwise from positive x-axis.

Solution:

$$G \frac{m_2 m_1}{r_{12}^2} = 3.75 \times 10^{-11} \text{ N} \quad (F_{21})$$

$$F_{31} = G \frac{m_3 m_1}{r_{13}^2} = 6.67 \times 10^{-11} \text{ N}$$

$$F = \sqrt{F_{21}^2 + F_{31}^2} = 7.65 \times 10^{-11} \text{ N}$$

$$\tan \theta = \frac{F_{21}}{F_{31}}$$

$$\theta = 29.3^\circ \text{ to the +ve x-axis}$$

Example 3: Discovering the Inverse Square Relationship (Coletta, P140)

Before Newton, Kepler had made a study of the motion of planets about the sun and had devised three laws based on his observations of the planets about the sun.

Kepler's Third Law states that:

The square of the time of revolution of a planet (i.e. its periodic times T) about the sun is proportional to the cube of its mean distances (r) from it, i.e. $T^2 \propto r^3$.

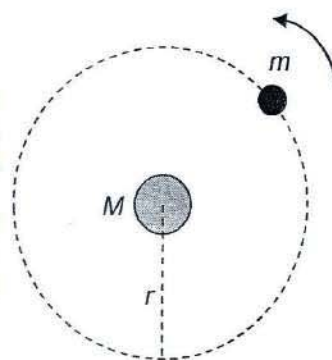
It is said that Newton could have discovered the inverse square relationship in the Law of Gravitation from Kepler's Third Law. Can you show that Kepler's 3rd Law reveals this inverse square relationship?

Solution:

Consider a planet of mass m revolving around the Sun of mass M in a circle of radius r .

As, the planet revolves, its velocity keeps changing as the direction of motion keeps changing. Hence, the planet is accelerating. From Newton's 2nd Law, we know that a resultant force must be providing this centripetal acceleration.

The gravitational attraction of the sun on the planet keeps the planet in orbit i.e. **the gravitational force acting on the planet (by the sun) provides for the centripetal force required for the planet to maintain circular orbit about the sun.**



By Newton's 2nd Law, $F_{\text{gravitational}} = ma_{\text{centripetal}}$

$$\Rightarrow F_{\text{gravitational}} = m(r\omega^2)$$

$$\Rightarrow F_{\text{gravitational}} = mr\left(\frac{2\pi}{T}\right)^2 = \frac{4\pi^2 mr}{T^2}$$

Kepler's 3rd Law states that: $T^2 \propto r^3 \Rightarrow T^2 = kr^3$, where k is a constant.

$$\Rightarrow F_{\text{gravitational}} = \frac{4\pi^2 mr}{T^2} = \frac{4\pi^2 mr}{kr^3} = \left(\frac{4\pi^2}{k}\right) \frac{m}{r^2}$$

$$\Rightarrow F_{\text{gravitational}} \propto \frac{m}{r^2}$$

Hence, the *gravitational force exerted by the sun on a planet* is proportional to the mass of the planet and *inversely proportional to the square of the distance between the planet and the sun.*

$\Rightarrow$ Obeys an inverse square relationship!

Example 4: How do you 'weigh' the Earth?

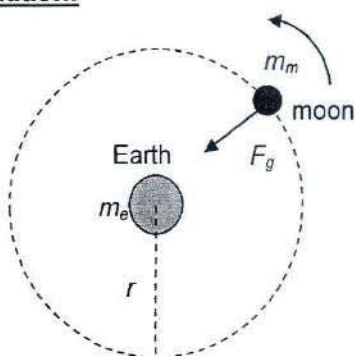
The Cavendish Experiment - Sixty Symbols



When Newton first discovered his Law of Gravitation, he had neither the value of G nor the mass of the Sun.

In 1798, a hundred and twenty-one years after Newton proposed the Universal Law of Gravitation, Henry Cavendish¹ conducted the first experiment to measure the force of gravity between masses in the laboratory, and obtained accurate values for the gravitational constant, G and the mass density of the Earth. With the value of G , it was possible to obtain a value for the mass of the Earth.

By considering the motion of the moon about the Earth, show that the mass of the Earth m_e is approximately 6×10^{24} kg. You may assume that distance between Earth and the moon $r = 4.0 \times 10^8$ m.

Solution:

The gravitational force acting on the moon by the Earth's gravitational field provides for the centripetal force required by the moon to move in orbit about the Earth.

By Newton's 2nd Law,

$$F_g = m_m a_c$$

$$\Rightarrow G \frac{m_e m_m}{r^2} = m_m r \omega_m^2 = m_m r \left(\frac{2\pi}{T_m} \right)^2$$

$$\Rightarrow m_e = \frac{4\pi^2 r^3}{G T_m^2}$$

Substituting for G , r_m and since we know that the moon takes approximately about 1 month to make one complete revolution around the Earth,

$$\Rightarrow m_e = \frac{4\pi^2 r^3}{G T_m^2} = \frac{4\pi^2 (4 \times 10^8 \text{ m})^3}{6.67 \times 10^{-11} \text{ N m}^2 \text{ kg}^{-2} (30 \times 24 \times 60 \times 60 \text{ s})^2} = 5.64 \times 10^{24} \text{ kg} \sim 6 \times 10^{24} \text{ kg}$$

Hence, we have found the mass of the Earth!

¹ https://en.wikipedia.org/wiki/Cavendish_experiment

7.1.1 Geostationary Satellites

Geostationary satellites are satellites with orbits in which the satellites are always positioned over the same geographical spot on Earth.² They appear stationary when seen from Earth.

Key Features of Geostationary Satellites:

1. **Period**
 - o 24 hours.
2. **Direction of rotation**
 - o same direction as the rotation of the Earth, i.e. from west to east.
3. **Plane of rotation**
 - o The plane that contains the orbit of the geostationary satellite as well as the equator.

The gravitational force on the satellite is always directed towards the centre of the Earth and so any circular orbit must have its centre at the centre of the Earth. If the orbit is not in the equatorial plane, the satellite will sometimes be over the northern hemisphere and sometimes over the southern hemisphere and so will not be geostationary.

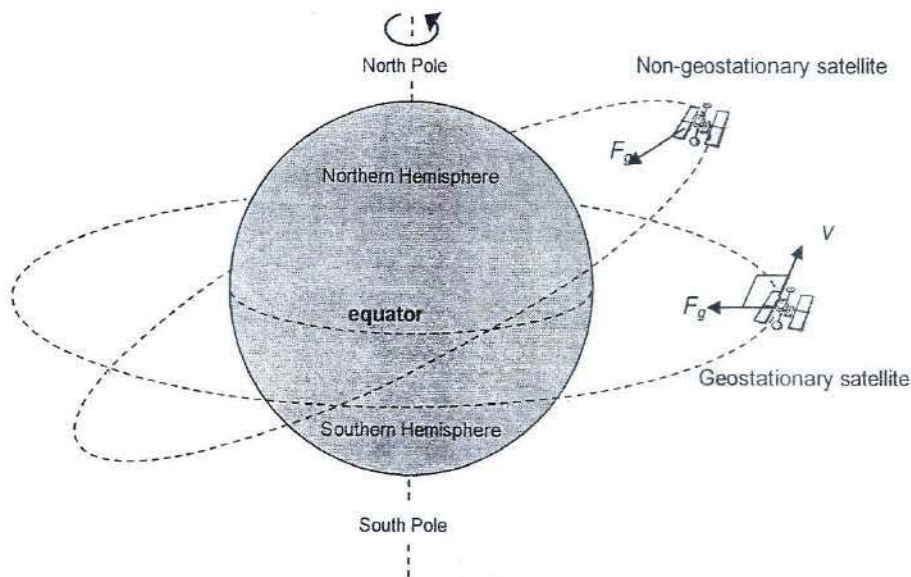


Figure not to scale

Advantages of Geostationary Satellites

1. As they remain stationary above the same point on Earth, they are ideal for use as communication satellites as they require no tracking to receive signals from Earth. This is why there is no need to keep adjusting the satellite dish to receive TV signals from a particular geostationary satellite.
2. As geostationary satellites are positioned at such a high altitude (see Example 5 for its height above the Earth's surface), they can view almost the whole Earth disk below them, rather than a small subsection. Hence, they are ideal for meteorological applications and remote imaging.

² Besides the geostationary orbits, satellites can have other orbits as well. Check out the following sites for different types of satellites, orbits and functions.

<http://earthobservatory.nasa.gov/Features/OrbitsCatalog/>

<http://www.satellites.spacesim.org/english/engineer/copy/index.html>

Disadvantages of Geostationary Satellites

1. As geostationary satellites are positioned at such a high altitude, the spatial resolution (i.e. the amount of detail shown) of their images tends to be not as good as satellites which are much closer to Earth.
2. Another problem with geostationary satellites is that since they are positioned above the equator, they cannot see the north or south poles and are of limited use for latitudes greater than 60-70 degrees north or south.

Example 5: Geostationary Satellite

A communications satellite of mass m is placed in a circular geostationary orbit.

Find its height above the Earth's surface.

(Take radius of the Earth to be 6.38×10^6 m and the mass of the Earth to be 5.98×10^{24} kg).

Solution:

$$\text{NZL: } F_g = ma_c$$

$$\begin{aligned} G \frac{Mm}{R^2} &= mR\omega^2 \\ &= mR \left(\frac{2\pi}{T} \right)^2 \\ &= mR \left(\frac{4\pi^2}{T^2} \right) \end{aligned}$$

$$\text{Re-arranging, } R^3 = \frac{GM}{4\pi^2} T^2$$

For geostationary satellites, $T = 24\text{h}$

$$R = \sqrt[3]{\frac{(6.67 \times 10^{-11})(5.98 \times 10^{24})(24 \times 60 \times 60)^2}{4\pi^2}}$$

$$= 4.23 \times 10^7 \text{ m}$$

Height above the earth's surface,

$$h = R - \text{radius of earth} = 4.23 \times 10^7 - 6.38 \times 10^6 = 3.57 \times 10^7$$

7.2 Gravitational Field Strength

The gravitational field is a region of space where a mass will experience a gravitational force.

Consider a particle of mass m placed at a point in a gravitational field (due to another mass M).



The particle would experience a **gravitational force**, F_g , in the gravitational field, given by:

$$F_g = mg$$

Definition: Gravitational Field Strength, g

The **gravitational field strength** g at a point is the gravitational force per unit mass acting on a small mass placed at the point.

$$g = \frac{F_g}{m}$$

where

g : gravitational field strength at that point
 F_g : gravitational force acting on mass m placed at that point in the gravitational field.

Notes:

1. S.I. unit of g : N kg^{-1} or m s^{-2} .
2. g is a vector quantity. Its direction is determined by the gravitational force acting on a mass m placed at the point in the gravitational field.
3. Mass m is a particle that is placed in the gravitational field. It is not the mass that is creating the gravitational field. Hence, mass m is often known as the *test* mass. The mass M that is creating the gravitational field is known as the *source* mass. The presence of the test mass is not necessary for the field to be present.
4. g is also commonly known as the *acceleration due to gravity* or *gravitational acceleration*.

7.2.1 Gravitational Field of a Point Mass

If a particle of mass M is placed at some point in space, it will set up a gravitational field around itself. Suppose we wish to determine the gravitational field strength g at a distance r away from this point mass M . We need to place a test mass m at the distance r away from this source mass M .



The test mass would experience a gravitational force F as it is in the gravitational field created by M .

By Newton's Law of Gravitation,

⇒ The magnitude of the gravitational force acting on m due to the gravitational field set up by M ,

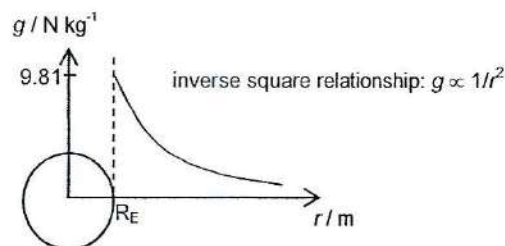
$$F = G \frac{Mm}{r^2}$$

From the definition of gravitational field strength,

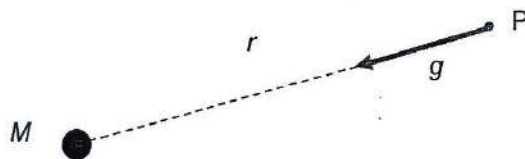
⇒ The magnitude of gravitational field strength due to M at a distance r away from M ,

$$g = \frac{\text{Magnitude of Gravitational force acting on test mass } m}{\text{Mass of test mass } m}$$

$$\begin{aligned} &= \frac{F}{m} \\ &= \left(G \frac{Mm}{r^2} \right) \\ &= \frac{m}{m} \\ &= G \frac{M}{r^2} \end{aligned}$$



Gravitational Field Strength due to a Point Mass



The magnitude of the gravitational field strength at a point due to a particle of mass M at a distance r away from the mass is given by:

$$g = G \frac{M}{r^2}$$

Note:

The equation is also valid for gravitational field strength outside spherical body of uniform density or spherical shells of uniform density, where its whole mass is concentrated at its geometric centre,³ r would then represent the distance between the test mass and the geometric centre of the source mass.

³ Read about Newton's Shell Theorem in Appendix A and gravitational field outside and within a sphere in Appendix B.

7.2.2 Representation of Gravitational Field

Besides using arrows to represent the gravitational field, another way to visualize the gravitational field in a region of space is to model it using field lines. The gravitational field lines are related to the gravitational field in the following manner:

1. **Direction**

The direction of the gravitational field strength g is tangent to the gravitational field line at each point in space.

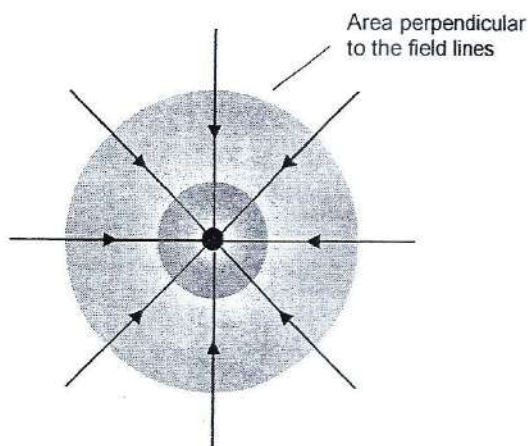
2. **Magnitude**

The magnitude of the gravitational field strength is proportional to the number of lines per unit area through a surface perpendicular to the lines.

Therefore, g is large when the lines are close together and small when they are far apart.

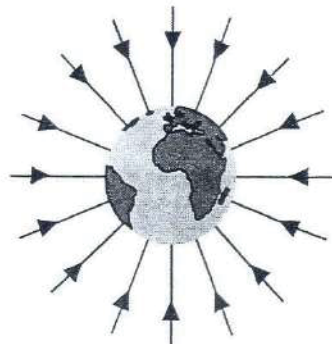
Gravitational Field Line Diagram

for a point mass

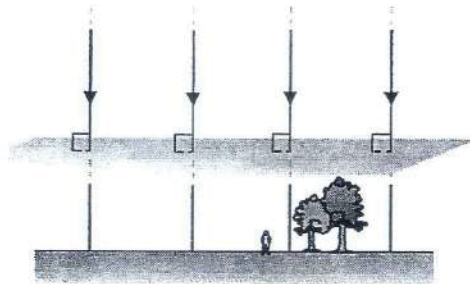


- The gravitational field in the vicinity of a point mass is always directed radially towards the point mass.
- The magnitude of the gravitational field is inversely proportional to the square of the distance from the point mass. Therefore, the number of field lines per unit area decreases the further away one is from the point mass, indicating the diminishing strength of the field.
- The field is really 3-dimensional, but on paper, we can only take a 2-dimensional slice of it. This is a radial or spherical field. Note that
 1. Field lines do not start or stop in empty space. They end on a mass and extend back all the way to infinity.
 2. Field lines never cross.

for Earth (spherical mass)



Close to the surface of the Earth, the field lines look like:



- The field lines are directed downwards (the direction in which a body near the Earth's surface would feel a gravitational force).
- The field lines are parallel and equidistant indicating that the field is constant, or uniform.
- The number of field lines per unit area is approximately constant for about 10 km in height from the Earth surface. Thus the field strength is approximately 9.81 m s^{-2} .⁴ (See Example 6 next page)

⁴ https://en.wikipedia.org/wiki/Gravity_of_Earth

Example 6: Gravitational Field Strength near the Surface of the Earth

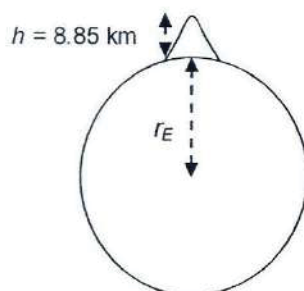
Assume that the Earth is not rotating and is a perfect sphere of uniform density.

- Calculate the gravitational field strength of the Earth at its surface.
- Given that Mount Everest is 8.85 km high, compute the gravitational field strength of Earth at the peak of Mount Everest.
- Comment on your values computed above.

You are given the following information:

- Mass of the Earth, $m_E = 5.98 \times 10^{24} \text{ kg}$
- Average radius of Earth, $r_E = 6.37 \times 10^6 \text{ m}$

Solution:



- The gravitational field strength at the surface of the Earth,

$$g_{\text{surface}} = G \frac{m_E}{r_E^2} = (6.67 \times 10^{-11} \text{ N m}^2 \text{ kg}^{-2}) \left[\frac{5.98 \times 10^{24} \text{ kg}}{(6.37 \times 10^6 \text{ m})^2} \right] = 9.83 \text{ m s}^{-2}.$$

- The gravitational field strength at the peak of Mount Everest,

$$g_{\text{everest}} = G \frac{m_E}{(r_E + h)^2} = (6.67 \times 10^{-11} \text{ N m}^2 \text{ kg}^{-2}) \left[\frac{5.98 \times 10^{24} \text{ kg}}{(6.37 \times 10^6 \text{ m} + 8.85 \times 10^3 \text{ m})^2} \right] = 9.80 \text{ m s}^{-2}$$

- Comparing the values computed in (i) and (ii), we see that

- $g_{\text{surface}} > g_{\text{everest}}$ by only approx. 0.03 m s^{-2} (about 0.3%).
- Hence, for the purpose of calculations, g may be taken to be constant near the Earth surface.

Furthermore, for heights h near the surface of the Earth such that $h \ll r_E$, the value of g may be taken to be constant.

7.2.3 Gravitational Field Strength (Acceleration due to Gravity)

If the only force acting on a mass m is the gravitational force, which is given by mg , where g is the gravitational field strength at the point, then applying Newton's second law,

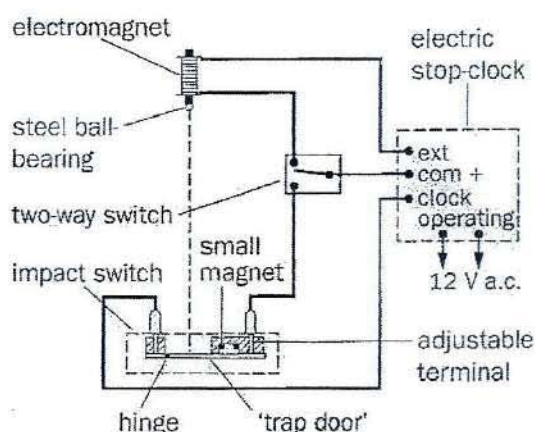
$$\begin{aligned} mg &= ma \\ a &= g \end{aligned}$$

Hence the mass will have an acceleration equal to the gravitational field strength g . Thus the gravitational field strength g is also known as *acceleration due to gravity*.

Note: There is a subtle difference between *acceleration due to gravity* and *acceleration of free fall*. Refer to Tutorial 7A D10(e).

7.2.4 Measurement of acceleration of free fall

One method of measuring the gravitational acceleration g is to examine the acceleration of an object when it is allowed to fall through a distance. Below is a description of the experiment.



In this experiment, a steel ball-bearing is made to fall a known distance, s , from rest and the time taken to fall this distance is measured using an electric stop-clock.

When the two-switch is changed to the 'down' position, the electromagnet releases the ball and simultaneously the clock starts. At the end of its fall, the ball opens the 'trap door' of the impact switch and the clock stops.

Assuming air resistance is negligible and applying the equations of motion,

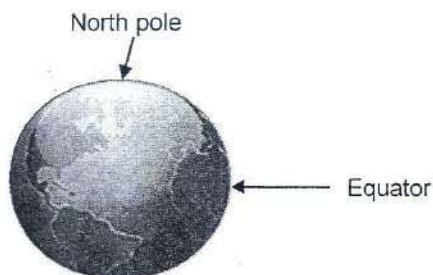
$$s = ut + \frac{1}{2}at^2 \Rightarrow g = \frac{2s}{t^2} \quad \text{where } t \text{ is the time taken to fall the distance } s.$$

7.2.5 Variation of g

However, when we try to measure the value of g at various regions on Earth, the value of g varies. At the poles, g is 9.83 m s^{-2} , while at the equator the value is 9.78 m s^{-2} . (In Singapore, it is about 9.79 m s^{-2}).

What accounts for this variation? There are mainly three factors affecting the value of g measured.

(1) The Earth is not a perfect sphere



Earth is approximately an ellipsoid, flattened at the poles and bulging at the equator. Its equatorial radius is greater than its polar radius by 21 km. Thus a point at the poles is closer to the dense core of the Earth than a point on the equator. This is one reason the gravitational field strength increases as one proceeds, at sea level, from the equator towards the poles.

(2) The density of the Earth is not uniform

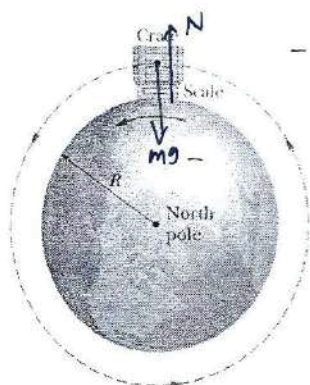
The density of the Earth varies from region to region over the Earth's surface. The northern hemisphere, for instance, has far more land area than the southern hemisphere. Accurate measurements of the gravitational pull are used to identify different densities of materials beneath the Earth's surface. These measurements are used in prospecting for oil, coal and other minerals.

(3) The Earth is rotating

The Earth rotates about an axis passing through the poles. Because of the Earth's rotation, the gravitational pull on a body at the equator has to also provide the body with a centripetal acceleration required for rotation. Therefore, the measured gravitational acceleration is smaller at the equator.

Example 7: How Earth's rotation affects the weight measured

Consider a crate of mass m resting on a weighing machine at the equator.



Take the Earth to be a uniform sphere of mass M and radius R , spinning on its axis.

- On the diagram, draw all the forces acting on the crate.
- Write down an expression for the reading of the weighing machine supporting this mass.
- Write down an expression for the 'apparent' gravitational acceleration indicated by the weighing machine.

Solution:

$$\text{By N2L, } F_g - N = ma_c$$

$$N = F_g - ma_c$$

$$\text{By N3L, mag. of } N' = \text{mag. of } N$$

$$N' = F_g - ma_c$$

$$c) g_{\text{apparent}} = \frac{W_{\text{apparent}}}{m} = \frac{N'}{m} = \frac{F_g - ma_c}{m} = \frac{F_g}{m} - a_c = \frac{F_g}{m} - r\omega^2$$

7.2.6 Weightlessness

There are two types of **weightlessness** – true weightlessness and apparent weightlessness.

True Weightlessness

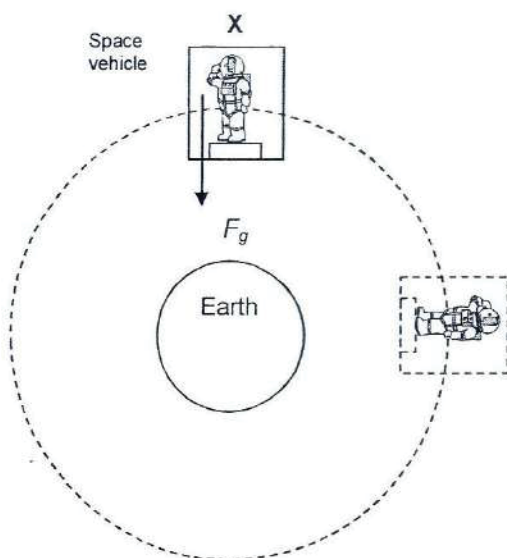
An object is **truly weightless** when there is **no net gravitational force acting on it**, such as when it is at an infinite distance away from any other body, or when it experiences opposing gravitational forces that exactly cancel one another out.

Apparent Weightlessness

An astronaut orbiting the Earth in a space vehicle with its thrusters off is often said to be 'weightless' as well. To understand why, let us try to 'weigh' this astronaut in the space vehicle.

Suppose an astronaut is standing on a weighing balance in a space vehicle orbiting the Earth.

Let us first consider the space vehicle, astronaut and weighing balance as a single system.



The only force acting on this system is the gravitational force due to the Earth's gravitational field, F_g .

Since the space vehicle is in a circular orbit around the Earth,
 $\Rightarrow F_g$ provides for the centripetal acceleration required for the space vehicle to move in a circular orbit.

By Newton's 2nd Law,

$$F_g = M_{sys} a$$

But

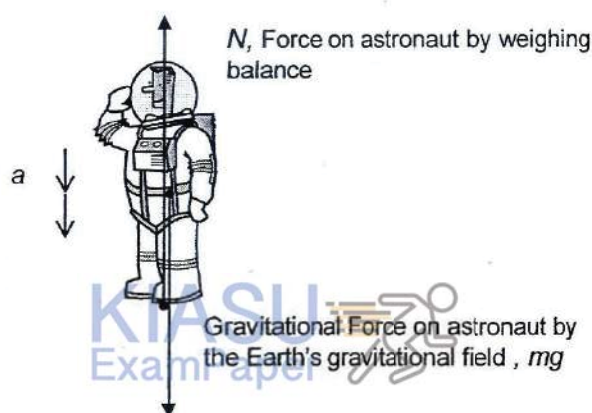
$$F_g = M_{sys} g$$

$$\Rightarrow M_{sys} g = M_{sys} a$$

Therefore, $a = g$

The space vehicle, the man and the weighing machine all have the same acceleration of g .

Now consider forces acting only on the astronaut at position X in the orbit



By Newton's 2nd Law,

$$mg - N = ma$$

where m is the mass of the astronaut and a is the acceleration of the astronaut.

However, from above, $a = g$.

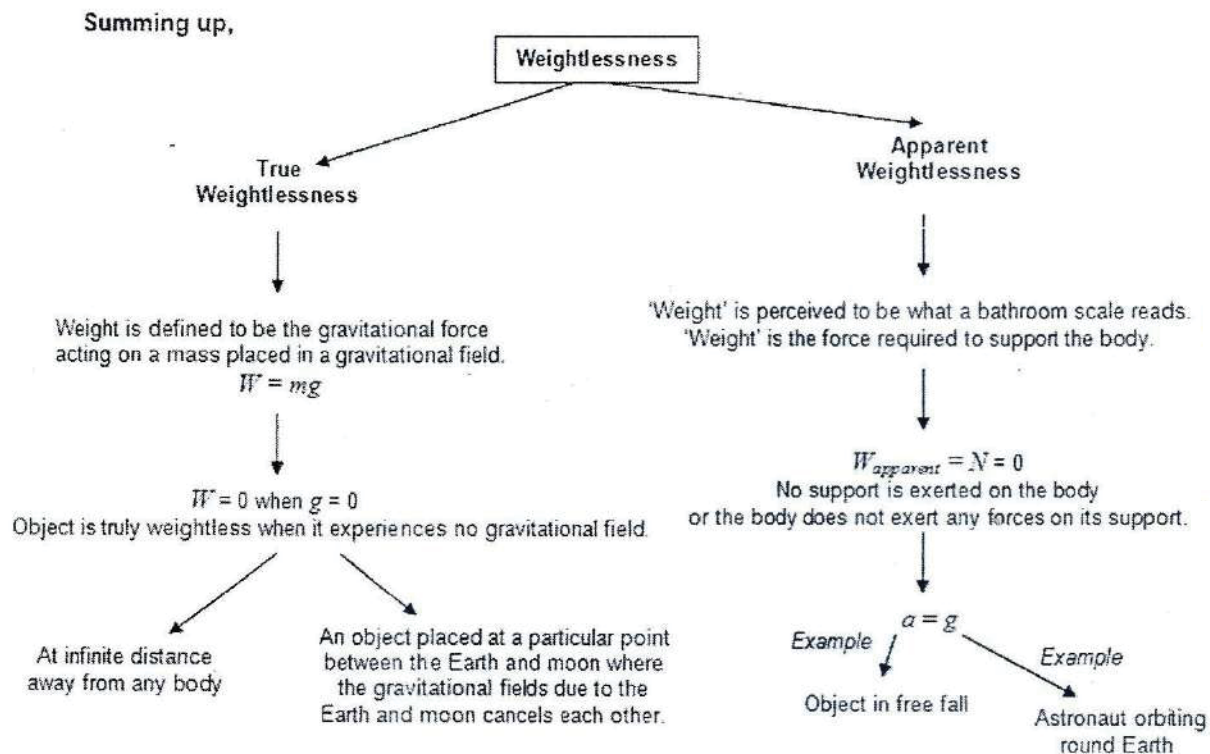
Therefore, $mg - N = mg$

$$\Rightarrow N = 0$$

Therefore, we see that the weighing balance does not exert a support force on the astronaut.

By Newton's 3rd Law, the astronaut does not exert a force on its support, the weighing machine. The reading indicated on the weighing machine is **zero**. The astronaut thus experiences the sensation of 'weightlessness'.

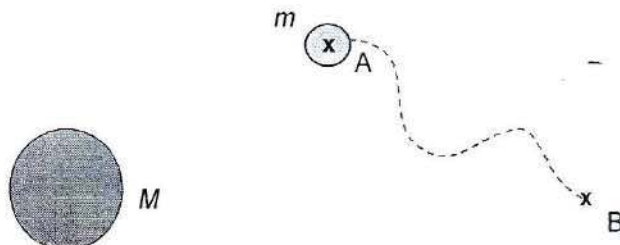
Hence, an object is considered to be (apparently) weightless when it exerts no contact force on its support. This happens when the object is accelerating with an acceleration of g , e.g. an object in free fall or an object orbiting around the Earth.



7.3 Gravitational Potential and Gravitational Potential Energy

7.3.1 Gravitational Potential Energy

Consider a region of space in which a gravitational field, exerted by source mass M , is present.



Consider a mass m placed in this gravitational field. If we apply an external force to this mass to move it between two points, from A to B, without any changes in its kinetic energy. By conservation of energy, whatever work is done by the external force would be equal to the change in gravitational potential energy of the mass m , i.e.

work done by an external force to move m from A to B in the gravitational field,

$$W = \Delta U_{AB} = U_B - U_A$$

where ΔU_{AB} : Change in gravitational potential energy of the particle as it moves from A to B

U_A : Gravitational potential energy at A

U_B : Gravitational potential energy at B

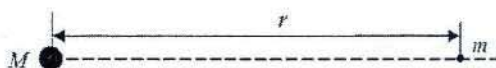
Choosing point A to be at an infinite distance from M , then the gravitational field strength due to M would be zero there. If we set the gravitational potential energy to be zero at that point ($U_A = U_\infty = 0$), then

$$W = \Delta U_{AB} = U_B - U_A = U_B$$

i.e. the work done by an external force to move a mass m from A (at infinity) to a point B without a change in kinetic energy is the gravitational potential energy of the mass at B.

Definition: Gravitational Potential Energy, U

The **gravitational potential energy** of a mass at a point in a gravitational field is the work done by an external force in bringing that mass from infinity to that point, without any change in kinetic energy.



Following the definition, the **gravitational potential energy U** of a system of two point masses M and m separated by a distance r is given by⁵ $\int_{\infty}^r \frac{GMm}{r^2} dr$. It can be shown that

$$U = -G \frac{Mm}{r}$$

⁵ A more rigorous derivation of the Gravitational Potential Energy (GPE) between 2 point masses and the gravitational potential of a point mass can be found in Appendix C. Although it is not in the syllabus, the derivation helps you associate to the definition of gravitational potential.

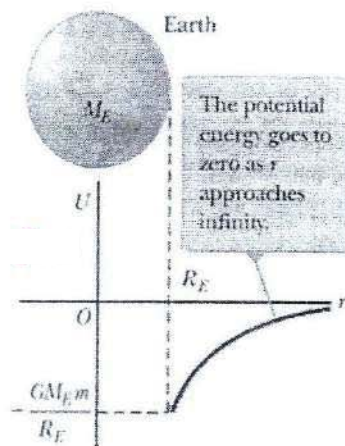
Note:

1. S.I. unit for U : joule (J).
2. The **gravitational potential energy is actually a property of the system of two masses** rather than of either mass alone. There is no way to divide this energy and say that so much belongs to one mass and so much to another. Therefore, it is more appropriate to say that it is the **gravitational potential energy of the system of M and m** .

Nevertheless, if $M \gg m$, as for the case of the Earth and the apple, we often use the phrase "the potential energy of the apple". This is because when an apple moves in the vicinity of Earth, changes in the potential energy of the Earth-apple system appear almost entirely as changes in kinetic energy of the apple, since the changes in the kinetic energy of the Earth is negligible.

3. U is a scalar quantity. Note that it can be positive or negative in value. In fact, gravitational potential energy is always negative. However, the negative sign has nothing to do with direction.
4. **Why is U negative and what is the significance of the negative sign?**

Gravitational force is attractive in nature. To bring a mass from infinity to the point in the field without changing its kinetic energy, an external force opposite to the gravitational force needs to be applied. Since the external force and displacement of the mass are in opposite directions, the external force does negative work on the mass in the process. As GPE at infinity is taken as zero, the GPE at any point must be negative.

**Example 8: Gravitational Potential Energy of a System of 3 Masses**

A system consists of three particles, each of mass m located at the corners of an equilateral triangle of side x . Determine the gravitational potential energy of the system.

Solution:

The total gravitational potential energy of the system is the sum over all pairs of particles. Gravitational potential energy obeys the superposition principle.

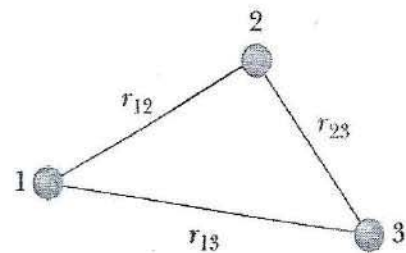
$$U_{\text{total}} = U_{12} + U_{13} + U_{23}$$

$$= -G \left(\frac{m_1 m_2}{r_{12}} + \frac{m_1 m_3}{r_{13}} + \frac{m_2 m_3}{r_{23}} \right)$$

Since $m_1 = m_2 = m_3 = m$ and $r_{12} = r_{23} = r_{13} = x$

the gravitational potential energy of the system $U_{\text{total}} = -3G \frac{m^2}{x}$

The absolute value of U_{total} represents the work needed to separate the particles by an infinite distance.



Example 9: Gravitational Potential Energy Changes near the Earth's Surface

Show that the expression $U = -G \frac{Mm}{r}$ can be reduced to the expression $mg\Delta h$ for calculations of changes in gravitational potential energy near the Earth's surface, i.e. for $h \ll$ radius of the Earth R_E .

Solution

At surface, GPE of mass,

$$U_i = -\frac{GMm}{R_E}$$

At height h above surface,

$$U_f = -\frac{GMm}{R_E + h}$$

Increase in GPE of mass,

$$\Delta U = U_f - U_i$$

$$\begin{aligned} &= -\frac{GMm}{R_E + h} - \left(-\frac{GMm}{R_E}\right) \\ &= GMm \left[\frac{1}{R_E} - \frac{1}{R_E + h} \right] \\ &= GMm \left[\frac{R_E + h - R_E}{R_E(R_E + h)} \right] \\ &= GMm \left(\frac{h}{R_E(R_E + h)} \right) \\ &\text{When } h \ll R_E, R_E(R_E + h) \approx R_E^2 \\ &\Delta U = GMm \left(\frac{h}{R_E^2} \right) = \frac{GM}{R_E^2} mh = mgh \text{ (shown)} \end{aligned}$$

Example 10: Escape Speed

The minimum speed with which a mass should be projected from the Earth's surface in order to escape Earth's gravitational field is known as the **escape speed** of the body from Earth.

Show that the escape speed of the mass m placed at the surface of the Earth is $\sqrt{2gR_E}$, given that the radius of the Earth is R_E and gravitational field strength at the Earth's surface is given by g .

Solution:

COE: Total energy of system at surface of earth = Total energy of system at infinity

$$GPE_s + KE_s = GPE_\infty + KE_\infty$$

$$-G \frac{Mm}{R_E} + \frac{1}{2}mv^2 = 0 + 0$$

$$V_{esc} = \sqrt{\frac{2GM}{R_E}} = \sqrt{2 \frac{GM}{R_E^2} R_E} = \sqrt{2gR_E} \text{ (shown)}$$

Assumptⁿ: • Earth has no atmosphere
• Earth is not spinning about its own axis

Example 11: Mechanical Energies of a Satellite in Orbit

A satellite of mass m orbits Earth of mass M in a circular path. The satellite has speed v and the radius of its orbit about Earth is R .

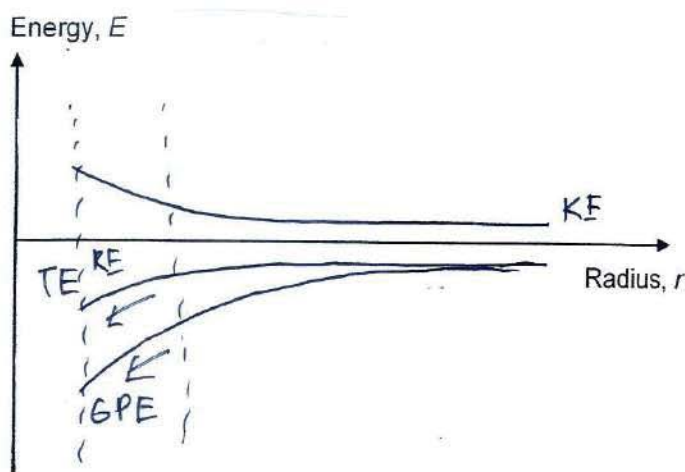
- Derive an expression for the KE and the total energy of the satellite in terms of G , M , m and R .
- Plot a graph showing how each of the following physical quantities associated with the satellite varies with distance from the centre of Earth:
 - kinetic energy;
 - gravitational potential energy;
 - total energy.

As the satellite orbits Earth, it loses energy due to air resistance.

- State whether the total energy E_T of the satellite becomes more or less negative.
- Hence state and explain the effect of this change on
 - the radius of the orbit; ↓
 - the speed of the satellite. ↑

TE ↓ by x, GPE ↓ by 2x, KE increase by x.

Solution:



a) $GPE = -\frac{GMm}{R}$

Gravitational force of earth on satellite F_g provides for centripetal force F_c for satellite to undergo uniform circular motion.

$$F_g = F_c \quad \left| \quad \frac{1}{2} \left(\frac{GMm}{R} \right) = \frac{1}{2} mv^2 \right| \quad \begin{aligned} TE &= GPE + KE \\ &= -\frac{GMm}{R} + \frac{GMm}{2R} \\ TE &= -\frac{GMm}{2R} \end{aligned}$$

$$\frac{GMm}{R} = mv^2$$

KIASU
ExamPaper

7.3.2 Gravitational Potential

Whenever a field is present in a region, we can always define a scalar property of the field at each point in the region. This property is called the *potential* of the field. This *potential* of a field is associated with the *potential energy per unit mass* that a test mass will possess when it is placed at the point in the field.

Definition: Gravitational Potential, ϕ

The **gravitational potential** at a point in a gravitational field is the work done per unit mass, by an external force, in bringing a small test mass from infinity to that point, without any change in kinetic energy.

Notes:

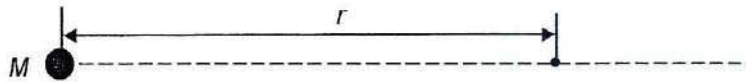
1. S.I. unit of gravitational potential ϕ : J kg^{-1} .
2. The gravitational potential is a **scalar quantity**.
3. The gravitational potential at infinity is taken to be zero.

Comparing the definition of **gravitational potential**, ϕ at a point in a gravitational field and **gravitational potential energy**, U of a mass, m placed at the same point in the gravitational field, we see that:

$$\phi = \frac{U}{m}$$

Together with the expression on p. 25: $U = -G \frac{Mm}{r}$, we deduce:

Gravitational Potential due to a Point Mass :

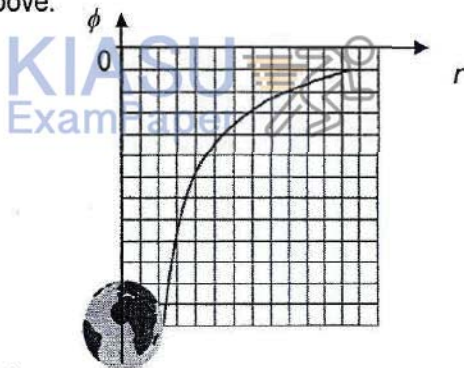


The gravitational potential ϕ at a point a distance r away from a source point mass M is given by:

$$\phi = -\frac{GM}{r}$$

Note:

1. The equation is also valid for any point outside masses of radial symmetry. This means that, for instance, we can assume the whole mass of the Earth to be concentrated at the Earth's centre. The gravitational potential *beyond the Earth's radius* can then be calculated using the formula given above.

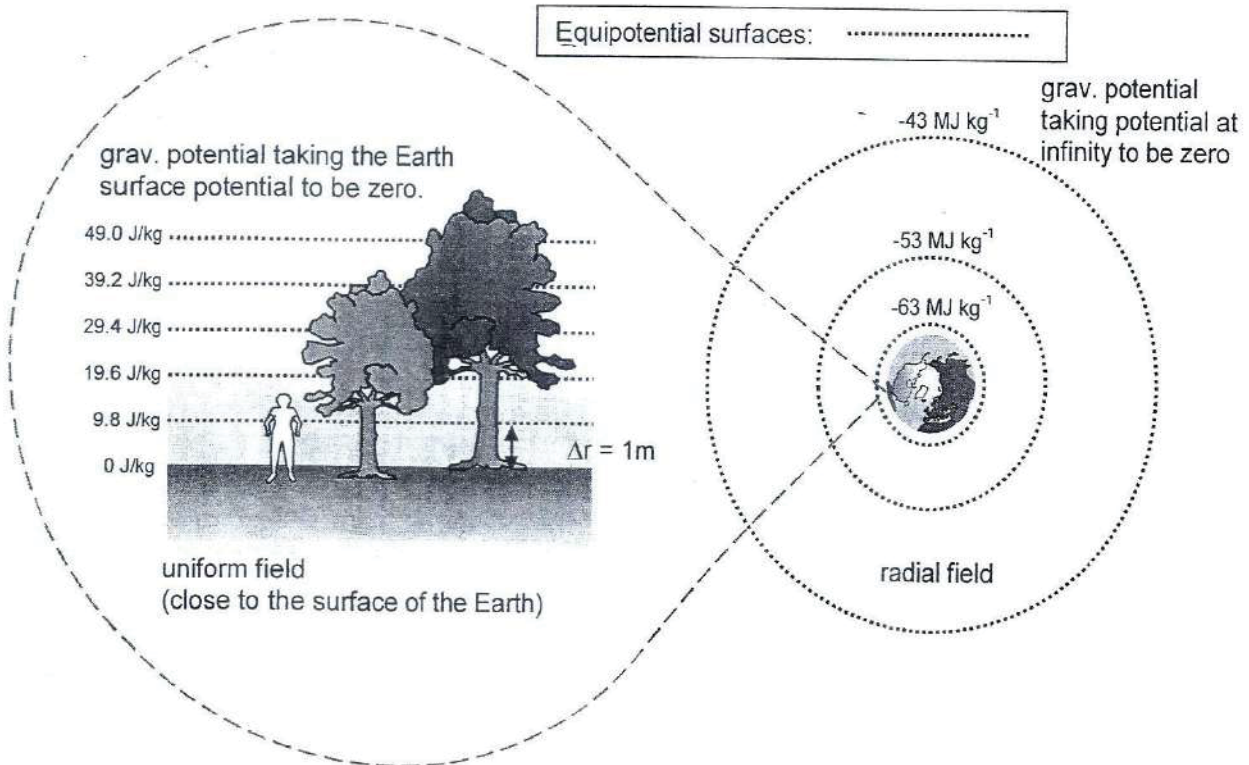


The gravitational potential at a point in the field is the work done per unit mass in bringing a test mass from infinity to that point.

Since the gravitational field is attractive and zero potential is at infinity, the gravitational potential is always negative.

2. All points at the same distance from the centre of the Earth will have the same gravitational potential. Points of the same potential are known as points of **equipotential**. We ignore the deviations from spherical symmetry due to mountains, seas and such.

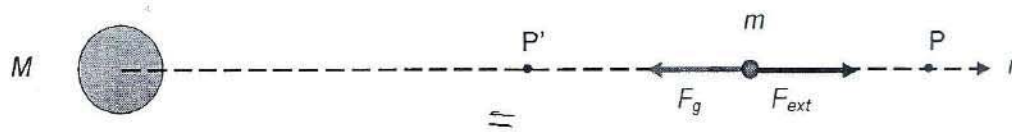
Equipotential lines/surfaces join points of equal potential. They are very simple in the cases of uniform fields (near the surface of the Earth) and radial fields (for point and spherical masses):



Equipotentials are always perpendicular to field lines. Diagrams of equipotential lines give us information about the gravitational field in much the same way as contour maps give us information about geographical heights (See next section).



7.3.3 Relationship between Potential and Field Strength



Consider a mass M fixed in space. Suppose we move a mass m from point P to point P' in the gravitational field of M by applying an external force F_{ext} such that there is no change in the kinetic energy of m . Then

$$\Rightarrow F_{ext} = -F_g$$

where F_g : gravitational force acting on m . (Note that F_{ext} and F_g vary with r .)

The change in gravitational potential energy of m , ΔU , is due to the work done W_{ext} by external force F_{ext} in moving a mass m from point P to point P' (without an increase in kinetic energy) i.e.,

$$\Rightarrow W_{ext} = \Delta U$$

$$\text{But } W_{ext} = \int_P^{P'} F_{ext} dr$$

$$\Rightarrow \int F_{ext} dr = \Delta U$$

$$\Rightarrow F_{ext} = \frac{dU}{dr}$$

$$\Rightarrow F_g = -\frac{dU}{dr}$$

Relationship between Gravitational Force F_g and Gravitational Potential Energy U

$$F_g = -\frac{dU}{dr}$$

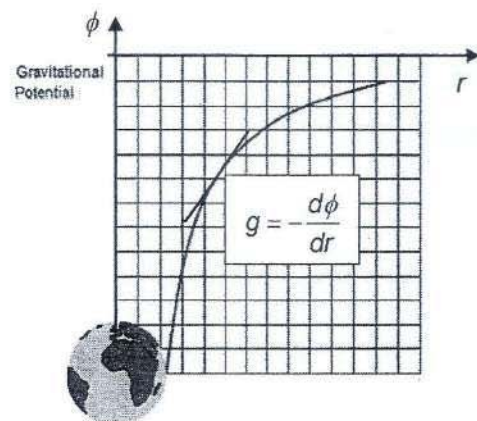
Dividing both sides of $F_g = -\frac{dU}{dr}$ by m , we obtain:

Relationship between Gravitational Field Strength g and Gravitational Potential ϕ

$$g = -\frac{d\phi}{dr}$$

Notes:

1. $\frac{d\phi}{dr}$ is known as the **potential gradient**.
2. The magnitude of the potential gradient at a point in the gravitational field gives the magnitude of the gravitational field strength at the point.
3. The gravitational field strength points in the direction of decreasing ϕ .



Example 12: Gravitational Potential between 2 Masses (Mod TYS Pg 91 Qn 38 N93/I/5; N2000/I/7)

The diagram (not to scale) represents the relative positions of the Earth and the Moon.



Figure 21

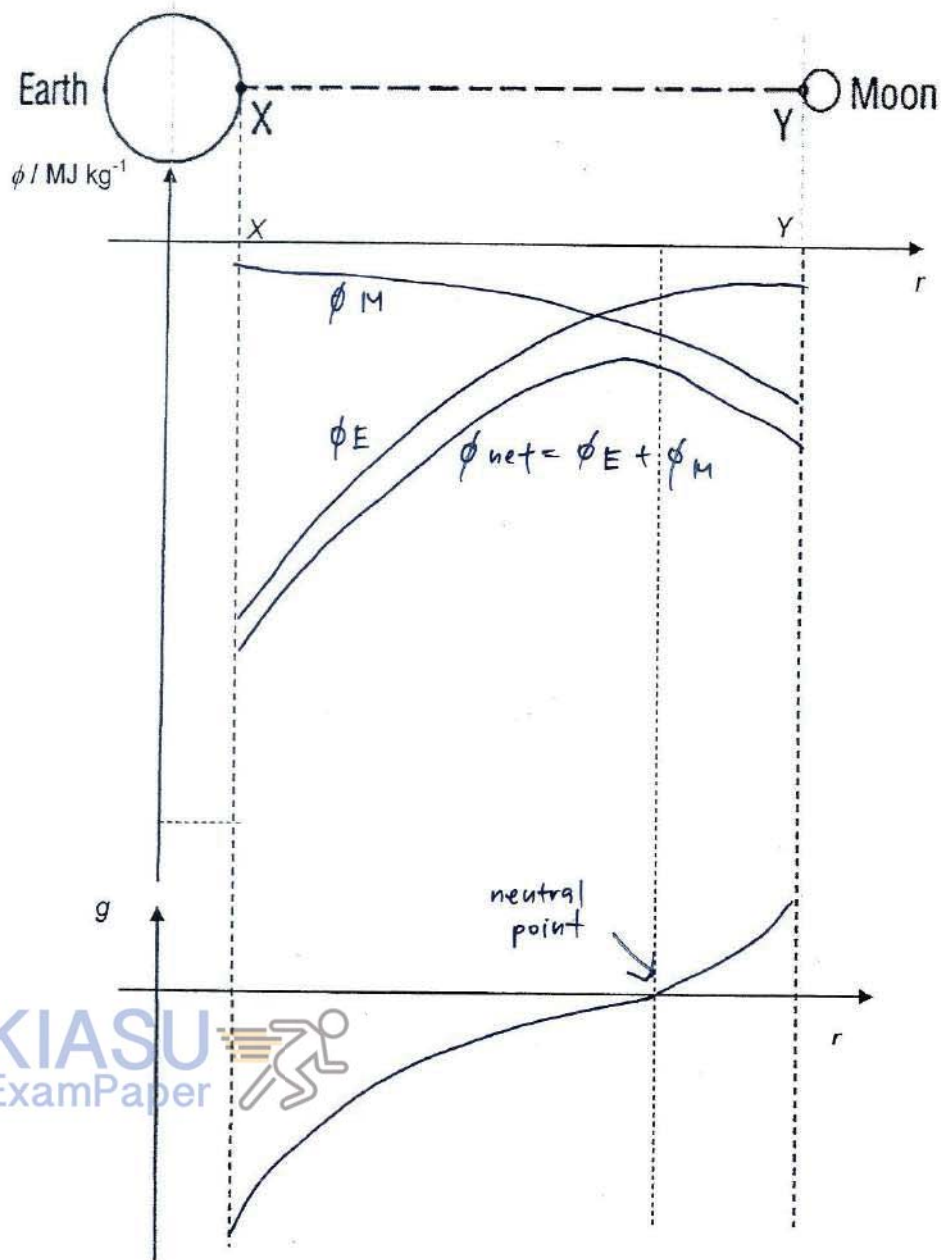
The line XY joins the surface of the Earth to the surface of the Moon.

(a) Sketch on the same axes graphs representing

- the gravitational potential due to the Earth between X and Y;
- the gravitational potential due to the Moon between X and Y;
- the resultant gravitational potential between X and Y.

(b) Sketch the corresponding variation of gravitational field strength with distance on the axis below.

Solution:



Appendices:

Appendix A: Gravitational Field due to a Uniform Spherical Shell⁶

(Adapted from Lieherherr, Martin. The Physics Teacher. Vol. 42. January 2004. "Gravitational Field Due to a Sphere: A Geometrical Argument." p5.)

From the statement of Newton's law of gravitation, we see that it applies only to point masses. This is a good approximation when the sizes of the objects are small compared to the distance between them, e.g. the Moon and the Earth.

Can Newton's law of gravitation be applied to an apple resting on the Earth's surface? From the point of view of the apple, the broad and level Earth, stretching out to the horizon beneath the apple, certainly does not look like a particle.

Newton solved this problem by using his law of universal gravitation and calculus to establish two theorems that apply to the gravitational force exerted by a thin spherical shell of uniform density.

Shell Theorem 1:

A uniform spherical shell of matter attracts a particle that is outside the shell as if all the shell's mass were concentrated at its centre.

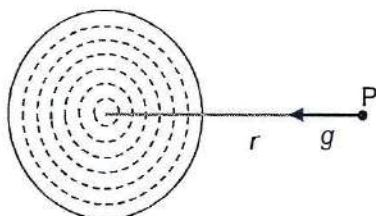
Shell Theorem 2:

A uniform shell of matter exerts no net gravitational force on a particle located inside it.

Let us now examine each shell theorem in more detail:

- **Shell Theorem 1:**

From Shell Theorem 1, we can see that any solid spherical object can be treated as many nested spherical thin shells with each of their centre of mass coinciding at the centre of the sphere.



The gravitational field at P due to each spherical shell is as if all the shell's mass were concentrated at the centre.

Since the centre of each spherical shell is common, the gravitational field at P, due to each shell, is directed towards the centre of the solid sphere.

Hence the magnitude of the resultant gravitational field at P can also be written as :

$$g = g_1 + g_2 + g_3 + \dots + g_N \quad \text{where } g_i : \text{gravitational field strength due to the } i\text{th shell.}$$

However, $g_i = G \frac{m_i}{r^2}$ where m_i : mass of the i th shell.

$$\begin{aligned} \text{Therefore, } g &= G \frac{m_1}{r^2} + G \frac{m_2}{r^2} + \dots + G \frac{m_N}{r^2} = \frac{G}{r^2} (m_1 + m_2 + \dots + m_N) \\ &= \frac{G}{r^2} (\text{mass of the sphere}) \end{aligned}$$

We can also conclude that

A sphere of uniform composition attracts a particle that is outside the sphere as if the entire sphere's mass were concentrated at its centre.

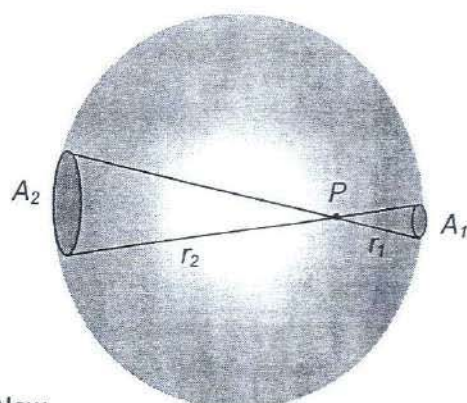
⁶ Proof of shell theorems requires techniques of integral calculus. If you are interested, refer to Resnick, Halliday and Krane. Physics -5th Edition. *Proof of Shell Theorems*. pp 306 – 307.

- **Shell Theorem 2:**

Although Newton proved both his theorems using calculus, we will present a simplistic geometrical proof below which makes use of the concepts of similarity.

Consider an arbitrary point P inside a thin spherical shell with homogeneous mass distribution.

Let σ be the mass per unit area of this spherical shell.



Let Point P be at the centre of a double cone which cuts out two circular areas A_1 and A_2 on the spherical shell.

We can therefore see that the gravitational field due to A_1 at P is directed towards centre of A_1 and the gravitational field due to A_2 at P is directed towards centre of A_2 , i.e. their directions are opposite to each other.

Now,

$$\frac{g \text{ due to } A_2}{g \text{ due to } A_1} = \frac{G \frac{\text{mass of } A_2}{r_2^2}}{G \frac{\text{mass of } A_1}{r_1^2}} = \frac{\frac{\sigma A_2}{r_2^2}}{\frac{\sigma A_1}{r_1^2}} = \frac{A_2 r_1^2}{A_1 r_2^2}$$

Since the two circular subtends the same angle, the cones are similar to each other.

$$\text{Therefore, } \frac{A_1}{A_2} = \left(\frac{r_1}{r_2}\right)^2 = \frac{r_1^2}{r_2^2}$$

$$\text{Hence, } \frac{g \text{ due to } A_2}{g \text{ due to } A_1} = \frac{A_2 r_1^2}{A_1 r_2^2} = \frac{r_2^2 r_1^2}{r_1^2 r_2^2} = 1,$$

i.e. the magnitude of the gravitational field due to A_1 = the magnitude of the gravitational field due to A_2 .

Since the gravitational field due to A_1 is equal in magnitude and opposite in direction to the gravitational field due to A_2 , the resultant gravitational field at P is zero.

Hence, any test mass placed at P will experience no gravitational force acting on it due to the spherical shell.

Furthermore, since P is arbitrarily chosen, we can also conclude that:

any mass placed within a uniform spherical shell will not experience any gravitational force due to the shell.

Appendix B: Variation of the Gravitational Field Strength above and below the Earth's Surface

Consider the Earth to be a non-rotating spherical planet of uniform density ρ . Take the radius of the Earth to be r_E , and the mass of the Earth to be m_E .

To determine the variation of g with respect to the distance from the centre of the Earth r , the following results (followed from Appendix A) are useful:

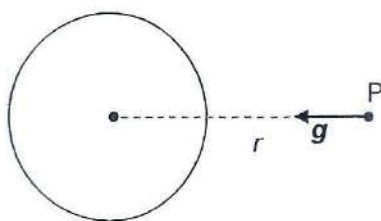
Result 1: The acceleration due to gravity **outside** a spherical body of uniform density is the same as if the entire mass of the body were concentrated at its centre.

Result 2: The acceleration due to gravity at all points **inside** a spherical shell of uniform density is zero.

With these results, we are able to obtain expressions for the acceleration due to gravity both above and below the Earth's surface.

Gravitational Field Strength above the Earth's surface ($r \geq r_E$)

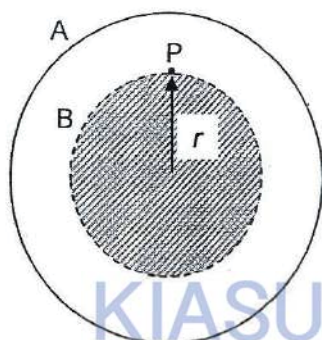
It follows from Result 1 that for $r \geq r_E$, $g = G \frac{m_E}{r^2}$; and at $r = r_E$, $g = G \frac{m_E}{r_E^2}$.



Since G and m_E are constants, $g \propto \frac{1}{r^2}$.

Gravitational Field Strength below the Earth's surface ($r \leq r_E$)⁷

Consider a point P below the Earth's surface and at a distance r away from the centre of the Earth.



The gravitational field at P,

$g = g$ at P due to spherical shell A + g at P due to spherical mass B

By Result 2,

Gravitation field at P due to shell A = 0, and

By Result 1,

Gravitation field at P due to spherical mass B

$$= G \frac{M_B}{r^2}$$

where M_B is the mass of the spherical mass B

⁷ The gravitational field strength below the Earth's surface is not explicitly required in the syllabus, it is only introduced here for completeness.

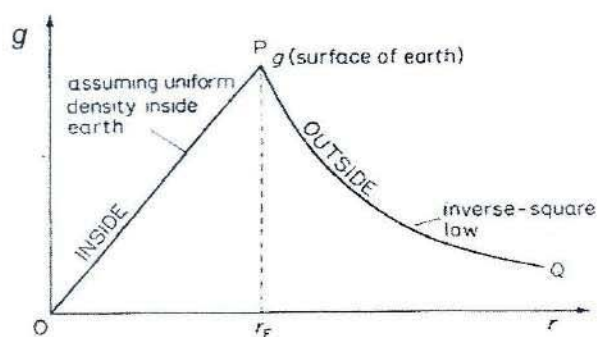
But $M_B = \rho$ (Volume of spherical mass B) $= \rho \left(\frac{4\pi r^3}{3} \right)$

Hence the gravitational field at P (with $r <$ radius of the Earth) is:

$$g = \frac{G}{r^2} \left(\frac{4\pi r^3 \rho}{3} \right) = \left(\frac{4\pi G \rho}{3} \right) r$$

Since $\frac{4\pi G \rho}{3}$ is a constant, we see that for $r < r_E$, $g \propto r$.

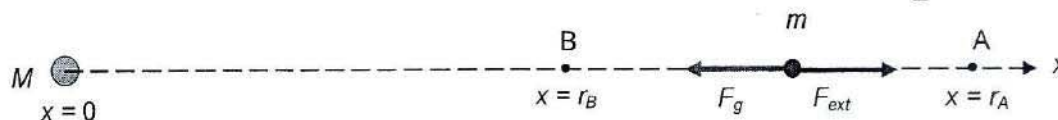
A sketch of g against r is shown in below.



Appendix C: Derivation of the Gravitational Potential Energy and Gravitational Potential

Consider a particle of mass m moving between two points A and B in the gravitational field exerted by another particle M . To find an expression for the gravitational potential energy of this system, we need to first find the work done by an external force in moving the mass m between the two points without a change in kinetic energy of m .

Suppose we set up the coordinate system as below.



By Newton's Law of Gravitation, the gravitational force⁸ F_g acting on the mass m at a distance x away from M ,

$$F_g = -G \frac{Mm}{x^2}$$

The negative sign here indicates that the gravitational force is pointing in the negative x direction.

For the kinetic energy of m to remain constant, we must move m at a **constant velocity** (acceleration = 0).

Therefore, the external force F_{ext} acting on m is opposite in direction and equal in magnitude to F_g .

$$\Rightarrow F_{ext} = G \frac{Mm}{x^2}$$

Since F_{ext} is not constant but varies with distance x away from M , to find the work done by this external force from moving the test mass m between A and B,

$$W = \int_{r_A}^{r_B} F_{ext} dx = \int_{r_A}^{r_B} G \frac{Mm}{x^2} dx = GMm \left[-\frac{1}{x} \right]_{r_A}^{r_B} = -GMm \left[\frac{1}{r_B} - \frac{1}{r_A} \right]$$

By conservation of energy, since there is no increase in the kinetic energy of the system,

$$W = \text{Change in gravitational potential energy of the system, } \Delta U$$

$$\text{Therefore, } W = \Delta U = U_B - U_A = -GMm \left[\frac{1}{r_B} - \frac{1}{r_A} \right] = \left[-G \frac{Mm}{r_B} \right] - \left[-G \frac{Mm}{r_A} \right]$$

Suppose we choose the reference point A for gravitational potential energy U to be zero at $r_A = \infty$, i.e.

$$U_A = U_\infty = \left[-G \frac{Mm}{r_A} \right] = 0$$

Then,

$$U_B - U_A = U_B = -G \frac{Mm}{r_B}$$

⁸ Note that gravitational force is an internal force to the system of M and m . To consider work done on m , we must always treat M and m as a single system, as m only experiences a force due to the gravitational field exerted by M .

Therefore, the gravitational potential of a test mass m placed at a distance of r away from the source mass M , is given by

$$U = -G \frac{Mm}{r}$$

From the definition of the potential energy at a point, the potential at a point at a distance r away from the source mass M ,

$$\phi = \frac{U}{m}$$

$$\Rightarrow \phi = -\frac{GM}{r}$$

Appendix D: Discovery of Neptune

In 1687, Sir Isaac Newton published *Principia*, which hypothesizes the inverse-square law of universal gravitation. In his own words, "I deduced that the forces which keep the planets in their orbs must be reciprocally as the squares of their distances from the centres about which they revolve; and thereby compared the force requisite to keep the Moon in her orb with the force of gravity at the surface of the Earth; and found them answer pretty nearly."

Newton's theory enjoyed its greatest success when it was used to predict the existence of Neptune based on motions of Uranus that could not be accounted by the actions of the other planets. We illustrate the situation schematically in the diagram on the right. The dominant interaction between Uranus and the Sun is indicated with the heavy line, but some perturbations associated with other masses are indicated by thin lines. Calculations by John Couch Adams and Urbain Le Verrier both predicted the general position of the new planet, and Le Verrier's calculations are what led Johann Gottfried Galle to the discovery of Neptune in 1846 within hours of turning his telescope on the region of the sky implicated by the calculations.



Years later, it was another discrepancy in a planet's orbit that showed Newton's theory to be inaccurate. By the end of the 19th century, it was known that the orbit of Mercury could not be accounted for entirely under Newtonian gravity, and all searches for another perturbing body (such as a planet orbiting the Sun even closer than Mercury) have been fruitless. This issue was resolved in 1915 by Albert Einstein's new general theory of relativity, which accounted for the discrepancy in Mercury's orbit.

Although Newton's theory has been superseded, most modern non-relativistic gravitational calculations still use it because it is much easier to work with and is sufficiently accurate for most applications.

(Adapted from https://en.wikipedia.org/wiki/History_of_gravitational_theory & <http://csep10.phys.utk.edu/astr161/lect/history/perturbations.html>)

Topic 8: Oscillations

H2 Physics C1 2016

Table of Contents

Learning Outcomes (LO).....	3
8.0 Introduction.....	4
What is periodic motion?.....	4
8.1 Simple Harmonic Motion (SHM).....	5
8.1.1 Terms used to describe SHM.....	6
8.1.2 SHM - displacement-time ($x-t$) relationship.....	7
8.1.3 SHM - velocity-time ($v-t$) relationship and acceleration-time ($a-t$) relationship.....	8
8.1.4 Definition of SHM – relationship between acceleration and displacement.....	10
8.1.5 Simple Harmonic Motion : velocity-displacement ($v-x$) relationship.....	13
8.2 Relationship between Simple Harmonic Motion and Uniform Circular Motion.....	15
8.3 Energy in SHM.....	17
8.3.1 Variation of Energy-displacement ($E-x$) graphs for a simple harmonic oscillator.....	17
8.3.2 Variation of Energy-time ($E-t$) graphs for a simple harmonic oscillator.....	19
8.4 Damped Oscillations, Forced Oscillations and Resonance.....	20
8.4.1 Free oscillations.....	20
8.4.2 Damped oscillations.....	21
8.4.3 Forced oscillations and resonance.....	23
8.4.4 Examples of destructive and useful resonance.....	26
Appendix 1 Period of a Simple Pendulum.....	28
Appendix 2 Period of Oscillating Mass on Vertical Spring.....	29
Appendix 3 Equations of motion for damped simple harmonic oscillations.....	30
Appendix 4 Demonstration of Forced Oscillation and Resonance - Barton's Pendulums.....	31

Oscillations

Candidates should be able to:

- (a) describe simple examples of free oscillations.
- (b) investigate the motion of an oscillator, using experimental and graphical methods.
- (c) understand and use the terms amplitude, period, frequency, angular frequency and phase difference and express the period in terms of both frequency and angular frequency.
- (d) recall and use the equation $a = -\omega^2 x$ as the defining equation of simple harmonic motion.
- (e) recognise and use $x = x_0 \sin \omega t$ as a solution to the equation $a = -\omega^2 x$.
- (f) recognise and use

$$v = v_0 \cos \omega t$$

$$v = \pm \omega \sqrt{x_0^2 - x^2}$$

- (g) describe, with graphical illustrations, the changes in displacement, velocity and acceleration during simple harmonic motion.
- (h) describe the interchange between kinetic energy and potential energy during simple harmonic motion.
- (i) describe practical examples of damped oscillations with particular reference to the effects of the degree of damping and the importance of critical damping in cases such as a car suspension system.
- (j) describe practical examples of forced oscillations and resonance.
- (k) describe graphically how the amplitude of a forced oscillation changes with frequency near to the natural frequency of the system, and understand qualitatively the factors which determine the frequency response and sharpness of the resonance.
- (l) show an appreciation that there are some circumstances in which resonance is useful and other circumstances in which resonance should be avoided.

RELEVANT E-LEARNING WEBSITES

- 1) "Vibration and Waves" free downloadable textbook <http://www.lightandmatter.com/area1book3.html>
- 2) Wikipedia: <http://en.wikipedia.org/wiki/Oscillation>, http://en.wikipedia.org/wiki/Simple_harmonic_motion
- 3) University of Salford Online Lessons: <http://www.acoustics.salford.ac.uk/feschools/waves/shm.htm#motion>
- 4) PHYSCLIP Online Lessons
 - a. Mechanics: SHM (http://www.animations.physics.unsw.edu.au/mechanics/chapter4_simpleharmonicmotion.html)
 - b. Waves and Sound: Oscillations (<http://www.animations.physics.unsw.edu.au/waves-sound/oscillations/>)
- 5) Useful physics interactive applets <http://ngsir.netfirms.com/englishVersion.htm#mechanics>
- 6) Interactive tutorial <http://www.physics.uoguelph.ca/tutorials/shm/Q.shm.html>



8.0 Introduction

Oscillations occur when a system is disturbed from a position of stable equilibrium. This displacement from equilibrium changes periodically over time. Thus, oscillations are said to be periodic and display periodic motion. Oscillations are very common in everyday life with familiar examples such as the motion of a clock pendulum or the vibrations of strings on musical instruments. Even atoms in a lattice vibrate. Oscillations are also important in many mechanical systems in the real world such as car suspension. It is thus very important to be able to understand these mechanical systems in order to control them in critical situations.

What is a periodic motion?

- **Periodic motion** refers to any motion that repeats itself at equal intervals of time.

Example 8.1: Periodic Motion

Which of the following cases are periodic oscillations?

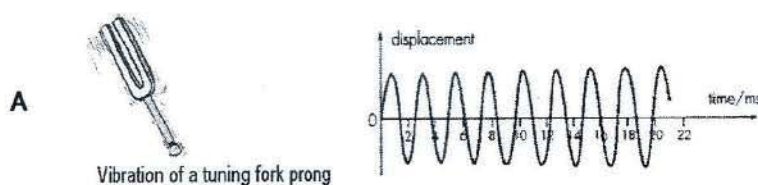


Fig. 8.1

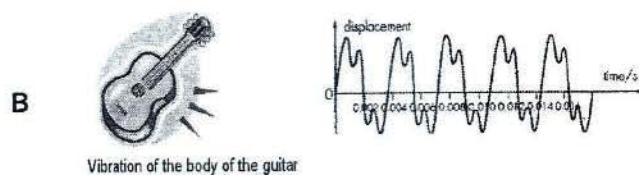


Fig. 8.2

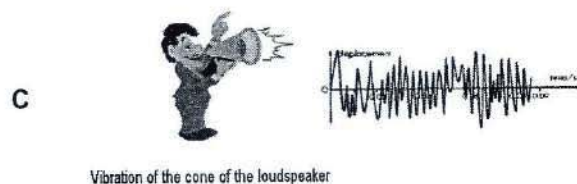


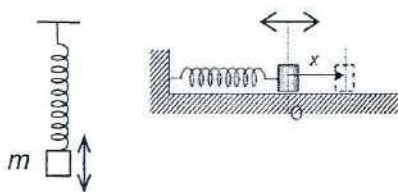
Fig. 8.3

8.1 Simple Harmonic Motion

For this topic, we are interested in **periodic oscillations** – in which an object moves about a fixed point such that the time it takes to complete one to-and-fro motion is the same.

One of the simplest oscillations we shall concern ourselves with is known as the **simple harmonic motion (SHM)**. Common examples of SHM include oscillations of a **simple pendulum** and a **spring-mass system** (assuming no loss of energy, so that the oscillations go on forever).

SHM provides a basis for the characterization of more complicated motions.

Simple Pendulum	Spring-Mass System
 $T = 2\pi\sqrt{\frac{l}{g}}$	 $T = 2\pi\sqrt{\frac{m}{k}}$
where T = period of oscillation l = length of the string g = gravitational acceleration (Refer to Appendix 1 for proof of formula)	where T = period of oscillation m = mass of the oscillating mass k = spring constant

In general, SHM is *isochronous* (the period and frequency are independent of the amplitude).

Let us take a closer look at a spring-mass system and learn the terms that describe the motion.

To study the motion we can attach a pen to the mass in such a way that it tracks the motion on a long vertical sheet of graph paper that is pulled to the left at a constant rate behind the system. This produces a trace on the graph paper which shows us how the displacement of the mass from its equilibrium position changes with time.

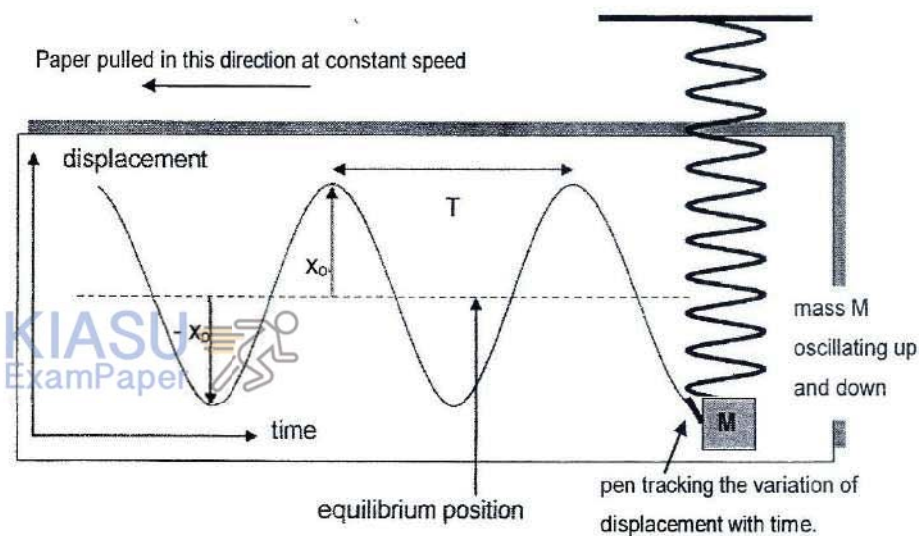


Fig. 8.4. Tracking the oscillation of a vertical spring-mass system

8.1.1 Terms used to describe SHM (with reference to Fig 8.4)

Terms	Definitions
Equilibrium ($x = 0$)	We often measure the displacement of the oscillator from the equilibrium position. Hence, the displacement $x = 0$ is designated at the equilibrium, i.e., the position at which the net force on the oscillator is zero .
Displacement, x	The displacement x is the linear distance of the oscillator from its equilibrium position ($x = 0$) in a specified direction. It is a vector; the $+$ / $-$ sign tells the direction of displacement from equilibrium. If you set the direction for displacement above equilibrium to be positive, then displacement below the equilibrium would be negative.
Amplitude, x_0	It is the magnitude of the maximum displacement of the oscillator from the equilibrium position. It is a scalar.
Period, T	It is the time taken for an oscillator to make one complete cycle of oscillation. S.I. unit: second.
Frequency, f	It is the number of cycles made by the oscillator per unit time: $f = \frac{1}{T}$ S.I. unit: hertz (Hz). $1 \text{ Hz} = 1 \text{ s}^{-1}$.
Angular frequency, ω	It is the rate of change of the phase angle with respect to time: $\omega = \frac{d\theta}{dt}$ Since frequency is a fixed value for an SHM oscillator, and a complete cycle of oscillation is represented by a phase angle of 2π rad, $\omega = \frac{2\pi}{T} = 2\pi f$ S.I. unit: rad s^{-1} . Note: this is also the familiar equation that we encountered in circular motion where ω is the angular velocity of an object moving in a circle.

8.1.2 SHM - Displacement-time (x - t) relationship

From Fig. 8.4, we see that the variation of the oscillations of a vertical spring-mass system follows a *sinusoidal*¹ pattern. All SHMs follow a sinusoidal motion.

Consider the spring-mass system in Fig. 8.4. Suppose at $t = 0$, the mass is at equilibrium ($x = 0$) and will next move in the positive direction, the displacement-time graph of the oscillatory system will look like:

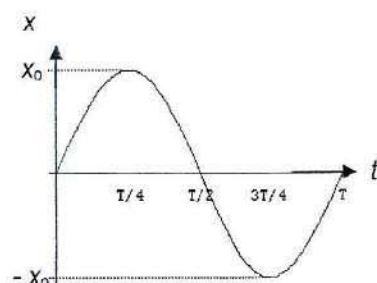


Fig. 8.5. A possible displacement-time graph of a simple harmonic oscillator

From the graph, we can see that the relationship between x and t is given by

$$x = x_0 \sin \omega t$$

where $\omega = \frac{2\pi}{T}$ is the angular frequency and x_0 is the amplitude of the oscillation.

Depending on the initial condition, i.e., when you start the timing, there are also other possible solutions. The following table lists some other possible solutions.

Initial conditions: (when $t = 0$)	Displacement-time (x - t) Graph	Displacement-time Equation
Oscillation starts from rest at amplitude ($v = 0$)		$x = x_0 \cos \omega t$ where x_0 is the amplitude.
Oscillation starts at positive displacement x' with initial positive velocity v		$x = x_0 \sin (\omega t + \phi_1)$ or $x = x_0 \cos (\omega t + \phi_2)$ where ϕ is the initial phase angle (not in syllabus)



Note: ωt is in radian. Remember to set your calculator to **RADIAN mode** before calculation!

¹ Sinusoidal is used to describe the shape of any sine or cosine curve regardless of whether it is shifted by some arbitrary horizontal amount.

8.1.3 SHM - Velocity-time (v - t) relationship and acceleration-time (a - t) relationship

From the topic of kinematics, we know that from the displacement-time relationship of an object, we can obtain its corresponding velocity-time and acceleration-time relationship by differentiating the equations or analysing the gradients of the tangents at the various points on the graphs accordingly (see Fig. 8.6).

We can similarly do so for SHM.

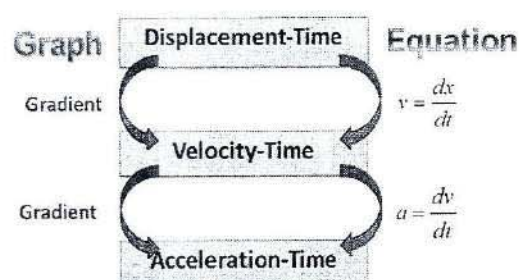


Fig. 8.6

$t = 0$ (initial conditions)	Example 1: $x = 0$, v is at maximum value	Example 2: $x = x_0$, $v = 0$
x-t graph		
Equation	$x = x_0 \sin \omega t$	$x = x_0 \cos \omega t$
v-t graph		
equation	$v = \omega x_0 \cos \omega t$	$v = -\omega x_0 \sin \omega t$
a-t graph		
equations	$a = -\omega^2 x_0 \sin \omega t$	$a = -\omega^2 x_0 \cos \omega t$

Given that $-1 \leq \sin \theta \leq 1$ and $-1 \leq \cos \theta \leq 1$, we can also see that:

- (1) Maximum value of velocity, v_0 :

$$v_0 = \omega x_0$$

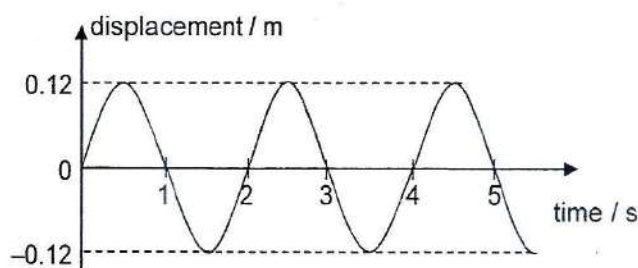
- (2) Maximum value of acceleration, a_0 :

$$a_0 = \omega^2 x_0$$

where ω is the angular frequency and x_0 is the amplitude of the oscillation.

Example 8.2 : Describing SHM in terms of time

The pendulum bob in a particular clock oscillates so that its displacement from a fixed point is as shown.



- (a) By taking the necessary readings from the graph, determine

- (i) the amplitude x_0 0.12
- (ii) the angular frequency, ω $\frac{2\pi}{2} = \pi$
- (iii) the maximum velocity v_0 and $v_0 = \omega x_0 = 0.38 \text{ m/s}$
- (iv) the maximum acceleration a_0 $\omega^2 x_0 = 1.2 \text{ m/s}^2$

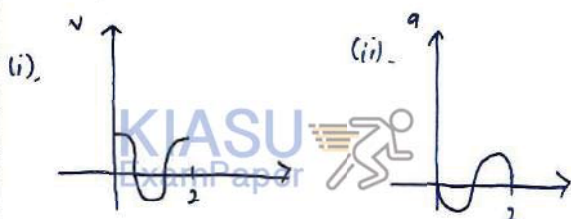
for this simple harmonic oscillator.

- (b) Hence, write down the equation that describes the displacement-time relationship. $x = 0.12 \sin 3.1t$

- (c) Sketch labelled graphs, illustrating the corresponding variation of the oscillator's

- (i) velocity,
- (ii) acceleration

with respect to time.



Example 8.3: Displacement-time Relationship for SHM

The displacement x (in m) of a body moving in SHM is given by the equation

$$x = 3.5 \sin 4.0t$$

- What is the amplitude of the motion?
- What is the angular frequency?
- What will be the displacement 0.20 s after oscillation has begun?
- Determine the velocity at time $t = 0.50$ s.

a) $x_0 = 3.5$

b) $\omega = 4.0 \text{ rad/s}$

c) $x = 3.5 \sin 4.0(0.20) = 2.5 \text{ m}$

d) $v = 3.5(4.0) \cos 4.0t = 14 \cos 4.0t$

$t = 0.50 \text{ s}, v = 14 \cos 4.0(0.50) = -5.8 \text{ m/s}$

8.1.4 Definition of SHM – relationship between acceleration and displacement

Returning to the table on page 8 and comparing the x - t and a - t relationship for both the examples, we see that $a = -\omega^2 x$. This is the defining equation for SHM!

Definition of Simple Harmonic Motion

Simple harmonic motion is a periodic motion in which the **acceleration** of the oscillator

- is **directly proportional** to the **displacement from its equilibrium point**, and
- is always in **opposite direction to the displacement**.

$$a = -\omega^2 x$$

where ω is the angular frequency and x is the displacement from the equilibrium point.

Notes:

- The **negative sign** in the equation indicates that **the acceleration in SHM is always in the direction opposite to the displacement from the equilibrium point**; in other words, the acceleration is always directed towards the equilibrium point.

- This is a 2nd order differential equation. Since $a = \frac{d^2 x}{dt^2}$, we can write $\frac{d^2 x}{dt^2} = -\omega^2 x$.

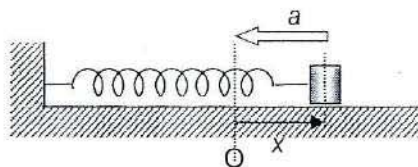
$x = x_0 \sin(\omega t + \phi)$ is a general solution of this differential equation, where ϕ depends on the starting position of the oscillator.

You are not required to know how to solve differential equations for the A-level Physics syllabus.

However, by differentiating the expression $x = x_0 \sin(\omega t + \phi)$ twice you can show that $\frac{d^2 x}{dt^2} = -\omega^2 x$.

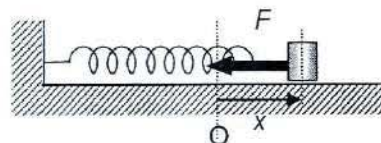
SHM of Spring-mass System Explained using Dynamics

Let us consider the oscillation of a mass oscillating at the end of a horizontal spring. Ideally, the mass of the spring is negligible, the mass slides without friction (undamped) on the horizontal surface and the spring obeys Hooke's Law. O is the equilibrium position, where the spring is unstretched.

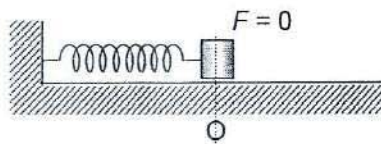


Let us look at how the restoring force of the spring on the mass varies with displacement x :

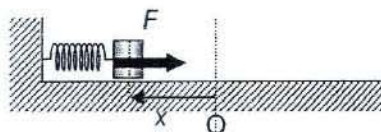
(a) when x is stretched to the right, the resultant restoring force is towards the left.



(b) when x is zero (unstretched), the resultant restoring force is zero.



(c) When x is compressed to the left, the resultant restoring force is towards the right.



Hence, we see that

1. the **resultant restoring force is always towards the equilibrium position**.
2. The **direction of this resultant force is always opposite to that of its displacement** at all points, except at equilibrium position, where $a = x = 0$.
3. According to Hooke's Law, $F = -kx$, where k is the spring constant. The restoring force, F , is proportional to the displacement, x , but directed opposite to it (as indicated by the negative sign).

By Newton's 2nd Law, $F_{net} = ma$

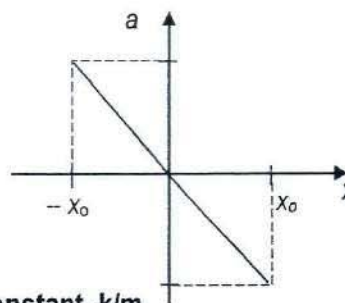
Since F_{net} is provided by the restoring force of the spring,

$$ma = -kx$$

$$a = -\frac{k}{m}x$$

Hence, the **acceleration is**

- **opposite in direction to the displacement and**
- **proportional to the displacement, with proportionality constant $-k/m$.**



which is the definition of SHM! In short, for mechanical oscillation to be possible, there must be a restoring force opposite to the displacement.

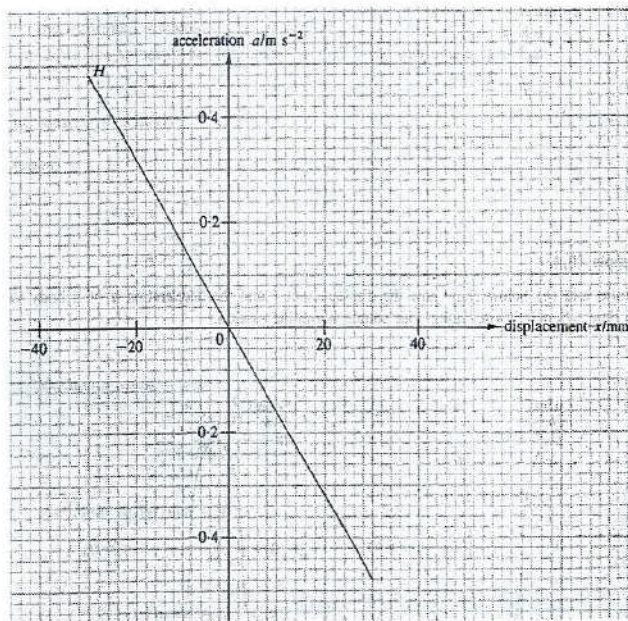
4. Comparing the expression above with the SHM equation $a = -\omega^2 x$: $\omega^2 = \frac{k}{m} \Rightarrow T = \frac{2\pi}{\omega} = 2\pi \sqrt{\frac{m}{k}}$

This is the expression we quoted for the period of the spring-mass system on page 5. Appendix 2 (pg. 29) provides the derivation of the period of a vertical spring-mass system, arriving at the same expression.

For a different SHM system, the angular frequency ω (and hence the natural frequency f and period T) will depend on different physical quantities, but the relationship $a = -\omega^2 x$ always holds.

Example 8.4: Relationship between acceleration and displacement for SHM

The graph represents the motion of a particle.



Find

- the amplitude of oscillation,
- the angular frequency of oscillation,
- the period of oscillation.

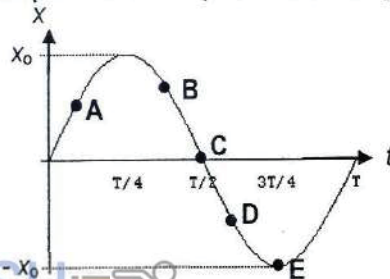
a) $x_0 = 30 \text{ mm}$

b) $-\omega^2 = g \text{ rad}^2/\text{s}^2$
 $\omega = 4.0 \text{ rad/s}$

c) $T = \frac{2\pi}{\omega} = 1.57 \text{ s}$

Example 8.5: Relationship between acceleration and displacement for SHM

For each box in the table below, indicate whether the direction of the displacement x , velocity v and acceleration a of each point on the displacement time graph is "+", "-" or "0".



	A	B	C	D	E
x	+	+	0	-	-
v	+	-	-	-	0
a	-	-	0	+	+

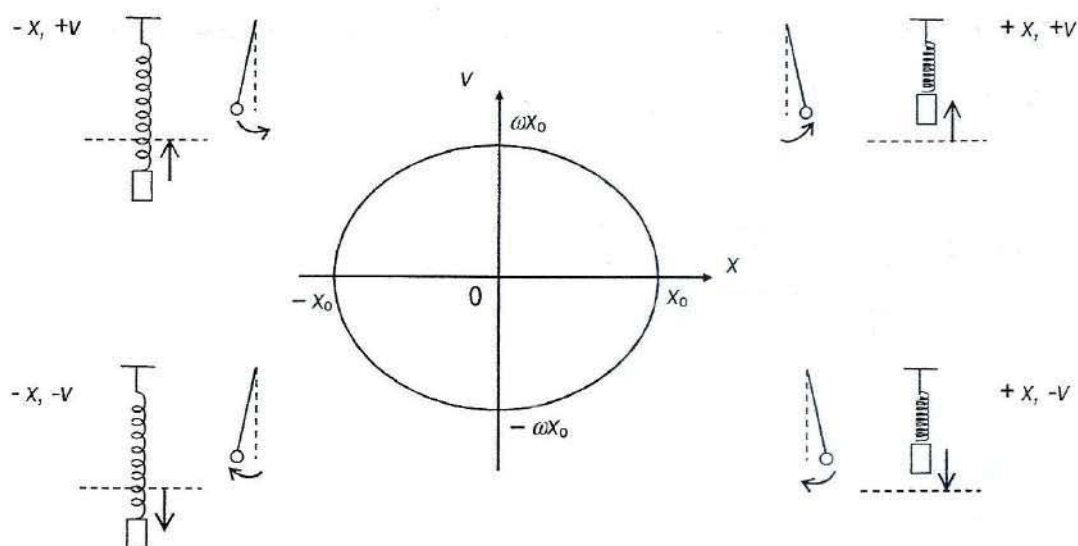
8.1.5 Simple Harmonic Motion: velocity-displacement (v - x) relationship

The velocity-displacement (v - x) relationship for SHM is given by

$$v = \pm \omega \sqrt{(x_0^2 - x^2)}$$

Note:

- 1) The equation is provided in the A-level Physics formulae list. It is applicable whichever position the oscillation starts from.
- 2) The “ $\pm$ ” sign implies that the body in SHM always passes through each position back and forth with the same speed.
- 3) At the amplitude positions, $v = 0$. At the equilibrium position ($x = 0$), the speed is maximum, $v_0 = \omega x_0$.
- 4) The v - x graph is an ellipse as shown.



Derivation of the relationship between velocity and displacement (not in syllabus)

Consider $x = x_0 \sin(\omega t + \phi)$ (1)

Differentiate (1) w.r.t. x : $v = \frac{dx}{dt} = \omega x_0 \cos(\omega t + \phi)$ (2)

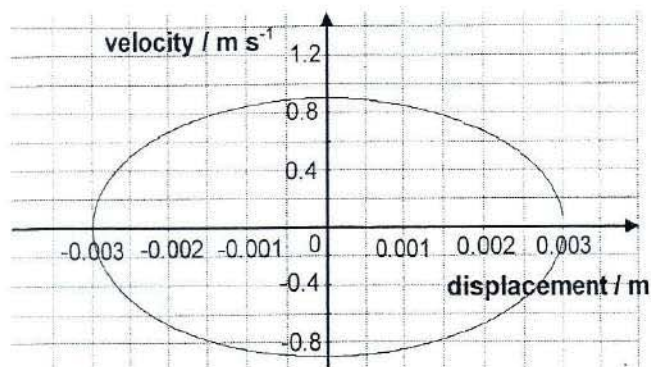
Given that $\sin^2 \theta + \cos^2 \theta = 1$, $\cos \theta = \pm \sqrt{1 - \sin^2 \theta}$

Substitute into equation (2): $v = \omega x_0 \left[\pm \sqrt{1 - \sin^2(\omega t + \phi)} \right]$



$$v = \pm \omega \sqrt{x_0^2 - x_0^2 \sin^2(\omega t + \phi)}$$

$$v = \pm \omega \sqrt{x_0^2 - x^2}$$

Example 8.6: Relationship between velocity and displacement for SHM

- Explain why there are two values of velocity for zero displacement.
- Explain why there are two values of displacement for zero velocity.
- Write down the equation which describes the graph.

(i). The object is momentarily at rest at both extreme ends

$$\begin{aligned} \text{(ii). } v_0 &= \omega x_0 = 0.9 & -x_0 &= -0.003 \\ -v_0 &= -\omega x_0 = -0.9 & x_0 &= 0.003 \end{aligned}$$

$$\begin{aligned} \text{(iii). } v &= \pm \omega \sqrt{x_0^2 - x^2} \\ &= \pm \frac{0.9}{0.003} \sqrt{0.003^2 - x^2} \end{aligned}$$

Example 8.7: Relationship between velocity and displacement for SHM

An object moving with simple harmonic motion has an amplitude of 0.020 m and a frequency of 20 Hz. Calculate

- the period of oscillation.
- the velocities at the equilibrium point and the position of maximum negative displacement.

Solution:

$$\text{a) } T = \frac{1}{f} = \frac{1}{20} = 0.050 \text{ s}$$

b) At equilibrium, $x = 0 \text{ m}$

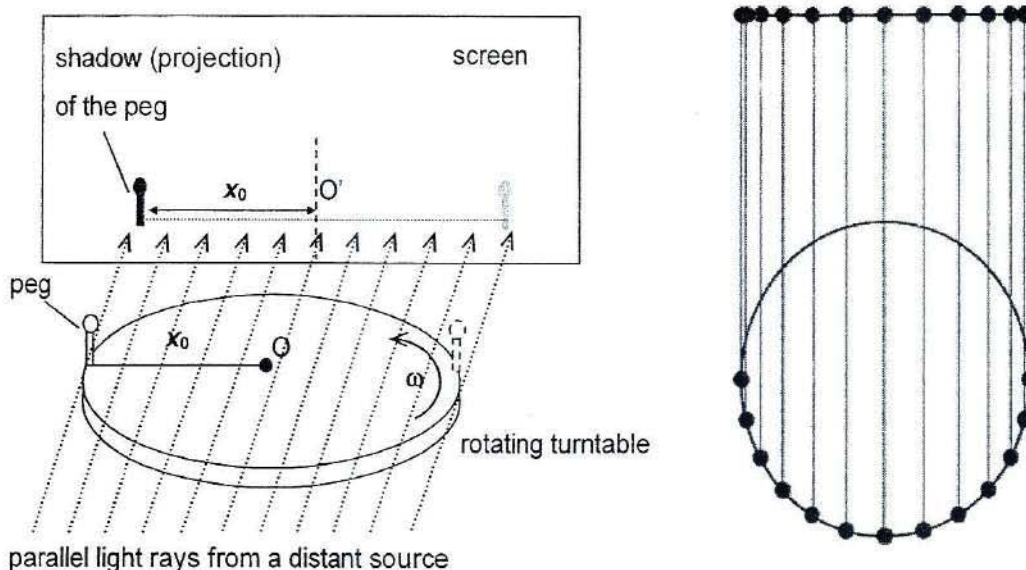
$$\begin{aligned} v &= \pm \omega \sqrt{x_0^2 - x^2} = \pm \omega x_0 = \pm 2\pi f x_0 \\ &= \pm 2\pi(20)(0.020) = \pm 2.51 \text{ m s}^{-1} \end{aligned}$$

At negative amplitude, $x = -0.020 \text{ m}$

$$v = \pm 2\pi(20)\sqrt{x_0^2 - x_0^2} = 0 \text{ m s}^{-1}$$

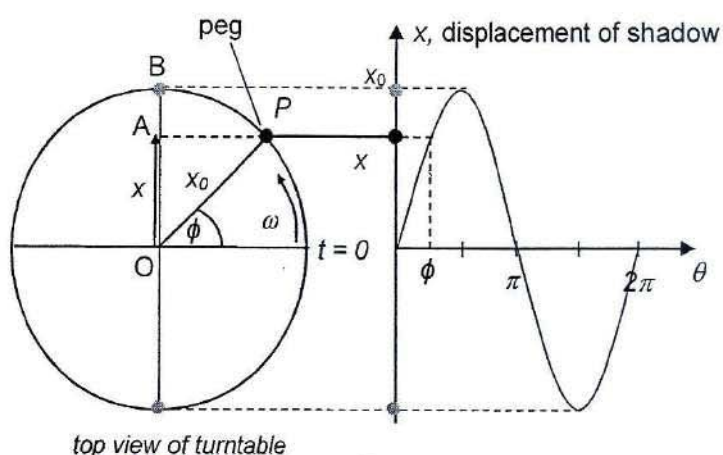
8.2 Relationship between Simple Harmonic Motion and Uniform Circular Motion

The experimental setup below illustrates how SHM is closely associated with circular motion. As the peg rotates on the turntable with constant angular velocity ω , its shadow on the screen moves back and forth in SHM.



Some observations

- 1) The motion is symmetrical about the fixed point, O' on the screen.
- 2) The time taken to complete one oscillation (period) is the same as the time taken to complete one cycle. Thus frequency of SHM = frequency of the circular motion.
- 3) The distance from the fixed point O' to the extremity (amplitude) is equal to the radius of the circle.
- 4) When the peg is shifted nearer to the centre of the turntable, the amplitude of the SHM decreases, but the period or frequency remains unchanged.
- 5) The shadow reaches maximum speed as it passes O' and is momentarily at rest at the extreme positions.
- 6) Variation of displacement of peg shadow with time



OA is the projection of OP (the displacement of the peg shadow) onto the vertical.

Applying trigonometry on $\triangle OAP$:

$$\Rightarrow x = x_0 \sin \phi \quad (OP = x_0)$$

Since $\phi = \omega t$, $x = x_0 \sin \omega t$ which is a typical SHM displacement-time equation.

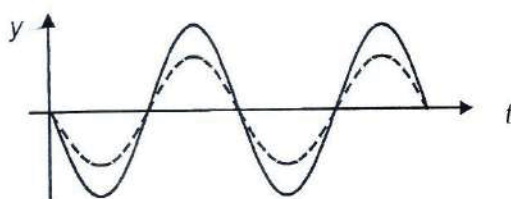
Phase & Phase Difference

In SHM, the angle ϕ is called the **phase (angle)** of oscillation, analogous to the angular displacement in circular motion.

Thus the angular frequency ω in simple harmonic motion is interpreted as the rate of change of phase angle of a simple harmonic oscillation.

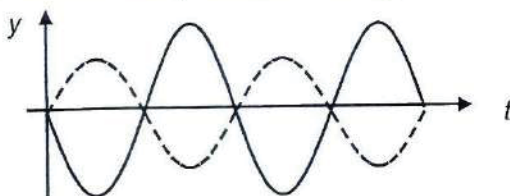
The idea of phase is useful for comparing two oscillators of the same frequency (and therefore period) that are not moving in step with each other due to a delay or advance at the start of the oscillations with respect to each other.

1. The difference in phase between two oscillations is called **phase difference $\Delta\phi$** .
2. Two oscillations are said to be **in phase** when they are in step, i.e., they reach their respective maximum displacements in the same direction at the same time. It does not matter what their respective amplitudes are.



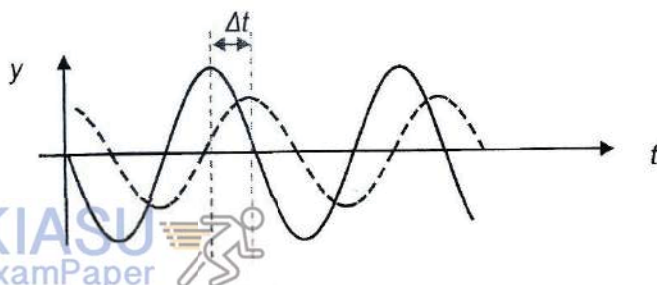
The phase difference $\Delta\phi = 0$ or any integer multiple of 2π rad.

3. Two oscillations are said to be **in anti-phase** when they are completely out of step, i.e., one oscillation reaches maximum displacement in a given direction at the same instant as the other oscillation reaches maximum displacement in the opposite direction.



The phase difference $\Delta\phi = \pi$ rad or any odd integer multiple of π rad.

4. Of course the two oscillations may not be in phase or in anti-phase but at some intermediate phase difference. In such a case, the phase difference can be calculated by the following: $\Delta\phi = \frac{\Delta t}{T} \times 2\pi$.



Access the following applet to learn more about phase difference between 2 oscillators.
<http://ngsir.netfirms.com/englishhtm/phase.htm>.

You will learn more about phase difference in the next chapter on Waves.

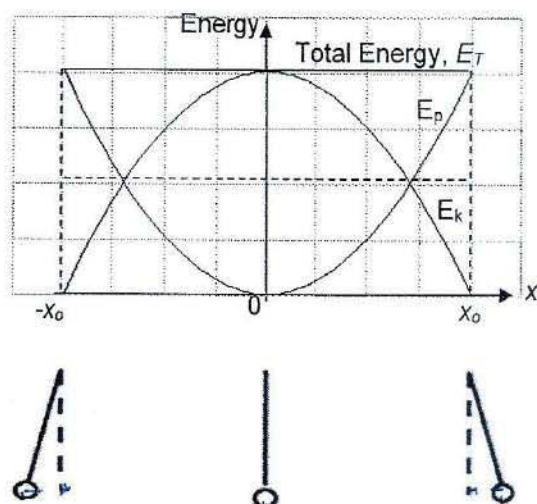
8.3 Energy in SHM

For a body moving in SHM, **the total mechanical energy E_T is conserved**. As the mass oscillates back and forth, the energy continuously transforms between potential energy E_p and kinetic energy E_k .

8.3.1 Variation of Energy-displacement (E - x) graphs for a simple harmonic oscillator

The sketch of the graphs can be deduced by observing the energy changes in a swinging pendulum or oscillating mass on a spring.

- As the oscillator passes through its equilibrium position, $x = 0$:
 $\Rightarrow$ its speed and hence its **kinetic energy** are **maximum**.
- At the extreme positions, $x = \pm x_0$,
 $\Rightarrow$ the speed and hence the **kinetic energy** are both **zero**.
 $\Rightarrow$ Assuming that there is no energy loss (e.g. no friction or air drag), the **total energy E_T of the oscillator must remain constant**.
 $\Rightarrow$ The **potential energy** will be **maximum**.



Note:

(1) **Variation of E_k with displacement x :**

For SHM, the velocity is given by $v = \pm \omega \sqrt{x_0^2 - x^2} \Rightarrow E_k = \frac{1}{2}mv^2 = \frac{1}{2}m\omega^2(x_0^2 - x^2)$

$$E_k = \frac{1}{2}m\omega^2(x_0^2 - x^2)$$

- $\Rightarrow$ Quadratic equation! Coefficient of x^2 is -ve.
- $\Rightarrow$ Graph is an inverted parabola (see figure above).

(2) **Variation of E_T with displacement x :**

- Total energy is constant. $\Rightarrow$ Graph of E_T vs. x is a horizontal straight line graph.
- Now, $E_T = \max E_K = \frac{1}{2} m \omega^2 x_0^2$ (when $x = 0$)

$$E_T = \frac{1}{2} m \omega^2 x_0^2$$

- The total energy E_T of SHM is proportional to the square of the amplitude, i.e., $E_T \propto x_0^2$

(3) **Variation of E_p with displacement x :**

Total Energy $E_T = E_p + E_K$

$$\Rightarrow E_p = E_T - E_K = \frac{1}{2} m \omega^2 x_0^2 - \frac{1}{2} m \omega^2 (x_0^2 - x^2) = \frac{1}{2} m \omega^2 x^2$$

$$E_p = \frac{1}{2} m \omega^2 x^2$$

- $\Rightarrow$ Quadratic equation! Coefficient of x^2 is +ve.
- $\Rightarrow$ Graph is a parabola (see figure above).

Example 8.8: Relationship between Energies and Displacement

A mass, $m = 0.50$ kg, connected to a light spring with spring constant $k = 20$ N m⁻¹ oscillates on a frictionless horizontal surface. Given that angular frequency $\omega = \sqrt{\frac{k}{m}}$, calculate the

- total energy of the system if the amplitude of the motion is 3.0 cm.
- kinetic energy of the system when the displacement is 2.0 cm and
 - potential energy of the system when the displacement is 2.0 cm.

$$\begin{aligned} \text{a) } E &= \frac{1}{2} m \omega^2 x_0^2 = \frac{1}{2} k x_0^2 \\ &= \frac{1}{2} (20)(0.030)^2 \\ &= 9.0 \times 10^{-3} \text{ J} \end{aligned}$$

$$\begin{aligned} \text{b) (i). } E_K &= \frac{1}{2} m v^2 = \frac{1}{2} m \omega^2 (x_0^2 - x^2) \\ &= \frac{1}{2} (0.50) \left(\frac{20}{0.50} \right) (0.03^2 - 0.02^2) \\ &= 5.0 \text{ mJ} \end{aligned}$$

$$\text{(ii). } E_p = E_T - E_K = 4.0 \text{ mJ}$$

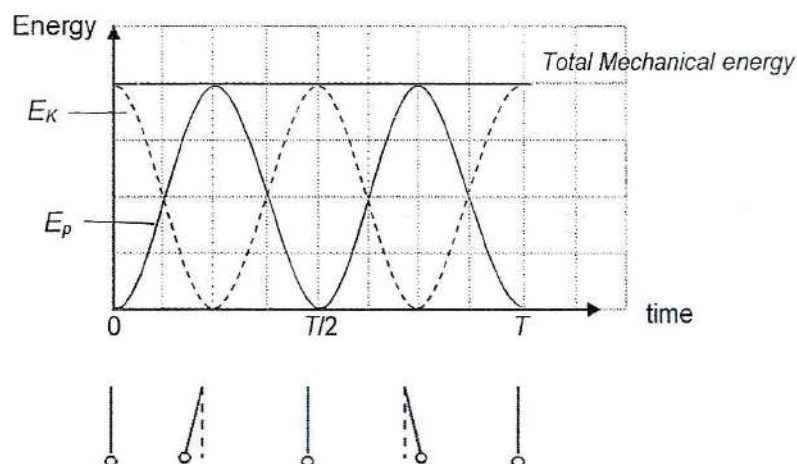
8.3.2 Variation of Energy with time graphs for a simple harmonic oscillator

The energy-time graphs depend on how the oscillation is started (initial conditions).

Again, without the actual equations, the shapes of the energy-time graphs can be deduced by observing a swinging pendulum or oscillating mass-spring.

Consider an oscillator started at its equilibrium position ($t = 0, x = 0, v = v_0$).

$$\Rightarrow x = x_0 \sin(\omega t) \text{ and } v = v_0 \cos(\omega t),$$



Note:

- (1) When the oscillatory motion has completed one cycle in one period, both kinetic and potential energies have gone through two cycles.

- (2) **Variation of total energy E_T with time t :**

○ Total energy is constant. $\Rightarrow$ Graph is constant at $E_T = \frac{1}{2}m\omega^2 x_0^2$

- (3) **Variation of kinetic energy E_K with time t :**

○ $E_K = \frac{1}{2}mv^2 = \left(\frac{1}{2}m x_0^2 \omega^2\right) \cos^2 \omega t$

$\Rightarrow$ Graph is a $\cos^2$ graph (it follows the shape of the v^2 - t graph).

- (4) **Variation of potential energy E_P with time t :**

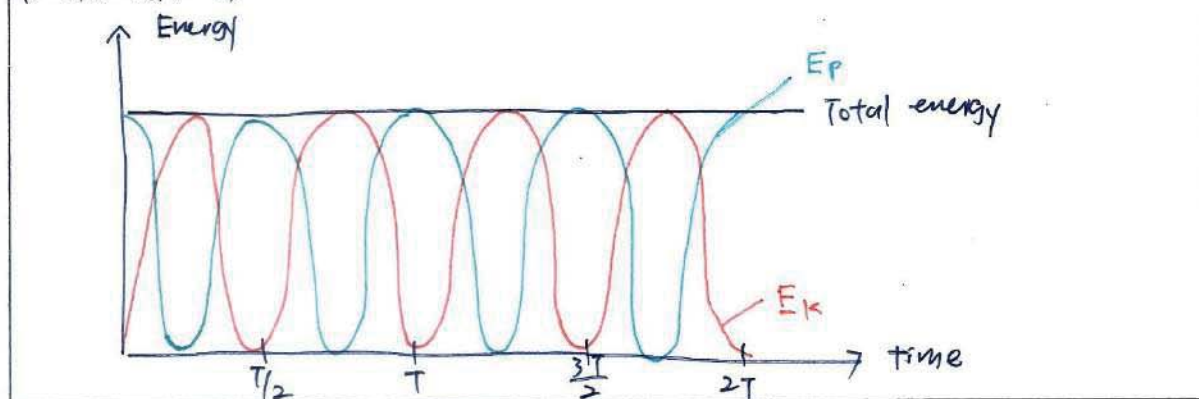
○ $E_P = \frac{1}{2}m\omega^2 x^2 = \left(\frac{1}{2}m\omega^2 x_0^2\right) \sin^2 \omega t$

$\Rightarrow$ Graph is a $\sin^2$ graph (it follows the shape of the x^2 - t graph).



Example 8.9: Relationships between Energies and time

Sketch the E_T - t , E_K - t , E_P - t graphs for an oscillator released from rest from its extreme position ($t = 0$, $x = x_0$, $v = 0$).

**8.4 Damped Oscillations, Forced Oscillations and Resonance****8.4.1 Free oscillations**

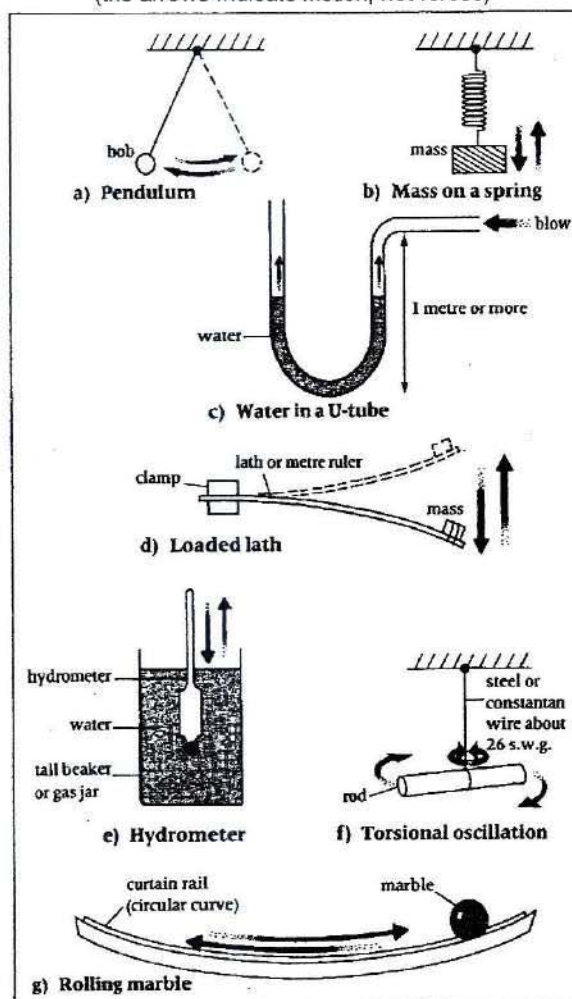
When a system is disturbed from its equilibrium position, such that it oscillates about this position with no resistive forces acting, the oscillation is said to be a **free oscillation**.

The frequency at which it oscillates is known as the **natural frequency** of the system. This frequency is characteristic of the system.

The simple harmonic motions that we have seen so far for the pendulum and spring-mass systems are ideal free oscillations in which no friction or any forms of resistance to motion exist. The oscillations would in theory continue forever.

Some examples of free oscillations

(the arrows indicate motion, not forces)



8.4.2 Damped oscillations

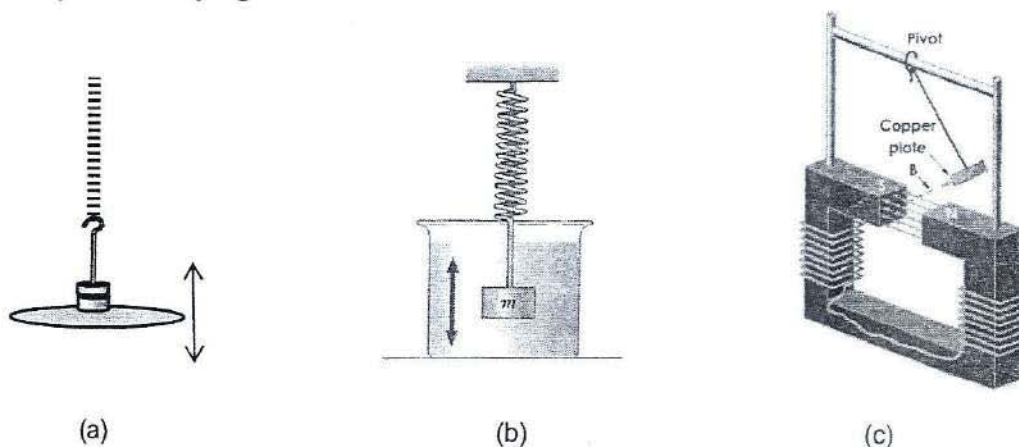
In an ideal oscillating system, there is no friction and the total mechanical energy is constant. The system in simple harmonic motion continues forever, with no decrease in amplitude.

Real-world systems experience inevitable friction and the mechanical energy of oscillations decreases with time due to the work done against friction. For instance, a simple pendulum gradually decreases in amplitude and stops eventually due to air resistance and friction at the support.

The progressive decrease in amplitude of any oscillatory motion caused by dissipative forces is called **damping**. Such an oscillatory motion is known as a **damped oscillation**. Due to damping, pendulum clocks require energy (wound up spring or batteries) to replenish energy loss in order for their amplitude to remain constant.

There are situations in which damping is introduced to reduce the amplitude of oscillation or bring the oscillation to rest.

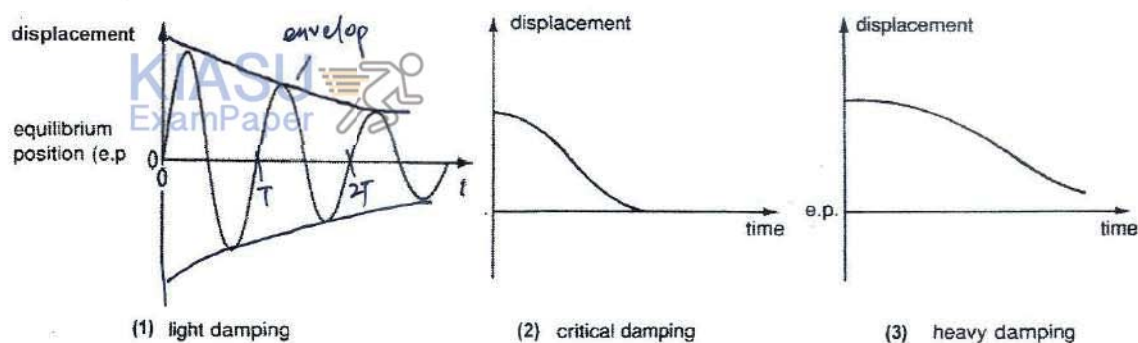
Commonly used damping methods:



- (a) Using air resistance, e.g., to bring an oscillating spring-mass system to rest faster, attach a piece of cardboard of negligible mass but large surface area to the base of the mass. The increase in air resistance increases damping during the oscillation.
- (b) Using fluid viscosity, e.g., damping can be increased by submersing the mass of a spring-mass system in a fluid that is more viscous than air, such as oil.
- (c) Using eddy currents, e.g., when a copper plate swinging freely passes through a magnetic field, eddy currents induced in the plate will oppose the motion and bring the plate to a stop (further discussed in Electromagnetic Induction in year 2).

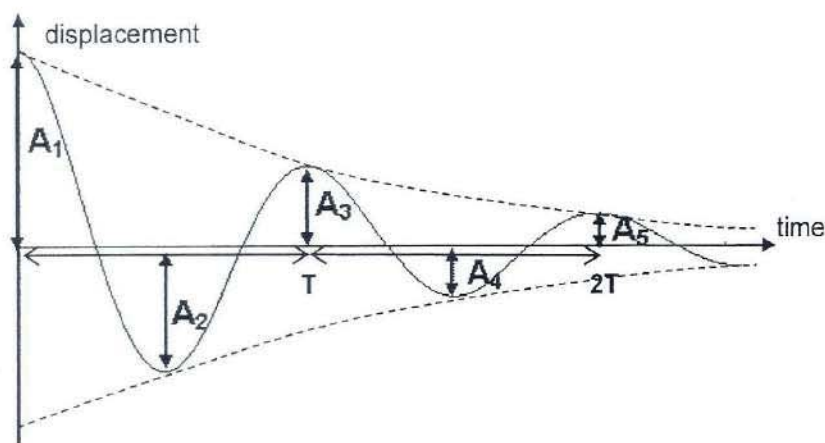
Extent of damping

Damping changes the behaviour of SHM systems.



(1) Light Damping (underdamped system)

- In a lightly damped system, the total energy of the system decreases with time as energy is dissipated when the system oscillates against resistive forces.
- The amplitude of the system decreases **exponentially with time**.

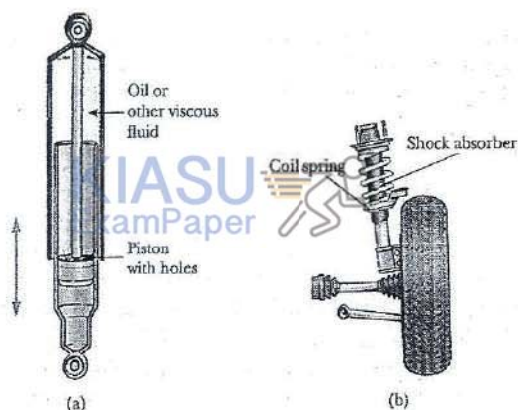


$$A_1 > A_2 > A_3 > A_4 > A_5$$

Note: It can be shown that the period in damped oscillations is slightly longer than the 'undamped' value by solving the differential equation involving the forces on the system. (This is not in the A-level syllabus – refer to Appendix 3 for details.)

(2) Critical Damping (critically damped system)

- If we are able to gradually increase the resistive force acting on an oscillatory system, damping increases till a point such that when displaced, the system **returns to the equilibrium position in the shortest possible time without oscillating at all**. The system is said to be **critically damped**.
- Critical damping is deliberately introduced into some systems to prevent continuous oscillations. For example, in analogue instruments such as balances, ammeters and voltmeters, the pointer is critically damped so that it stops in the shortest possible time to indicate the reading, instead of oscillating about the reading (underdamped) or take an unnecessarily long time to crawl to the reading (overdamped).
- Critical damping is also employed in car suspension systems to ensure a smooth ride when the car moves on a bumpy road. The shock absorbers in the suspension system have hydraulic pistons to prevent the car from bouncing up and down excessively after hitting a bump. The damping of a car can be tested by pushing down the car with your weight and then releasing it; the car should rapidly return to its normal height with negligible overshoot.



- (a) A shock absorber consists of a piston oscillating in a chamber filled with oil. As the piston oscillates, the oil is squeezed through holes between the piston and the chamber, causing a damping of the piston's oscillations.
- (b) One type of car suspension system, in which a shock absorber is placed inside a coiled spring at each wheel.

(3) Heavy Damping (overdamped system)

- When the damping force is increased beyond the point of critical damping, we say that the system is **overdamped** or that it experiences heavy damping. Once displaced from equilibrium, the oscillator takes a very long time to return to its equilibrium position.
- An example would be of certain doors which are so heavily damped that they take forever to close.

8.4.3 Forced oscillations and resonance**Forced oscillations**

As mentioned earlier, most real systems will experience some form of damping and will eventually stop oscillating. To keep them oscillating, one needs to periodically put in energy to the systems to make up for the energy lost. In other words, one needs to apply a periodic force to the systems.

- An oscillation under the influence of an external periodic force is called a **forced oscillation**.
- The external force that is used to sustain (or drive) the oscillations is known as the **driving force**.
- The frequency with which the periodic force is applied is called the **forced frequency** or **driver frequency**.

One example of forced oscillation is pushing a child on a swing at the playground; one needs to continually give the swing a push at the right moment to sustain its oscillation.

Resonance

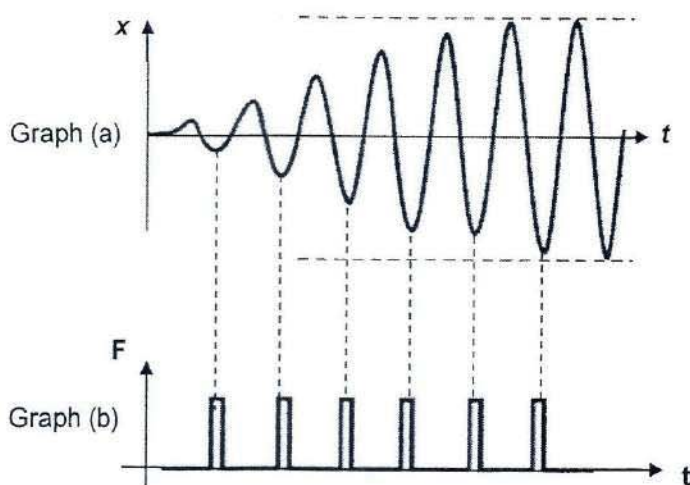
Consider the above example of a child on a swing. We have mentioned that because of damping, the oscillations will eventually die off. However, we can sustain the oscillations by periodically applying a driving force.

If the frequency of the driving force matches the natural frequency of the oscillatory motion of the child-swing system, an interesting phenomenon happens - the final amplitude of the oscillation would be the greatest compared to the amplitudes of oscillation for other frequencies of driving force. This phenomenon is known as **resonance**.

Resonance is a phenomenon in which an oscillatory system responds with maximum amplitude to an external periodic driving force, when the frequency of the driving force (driver frequency) equals the natural frequency of the driven system.

Note:

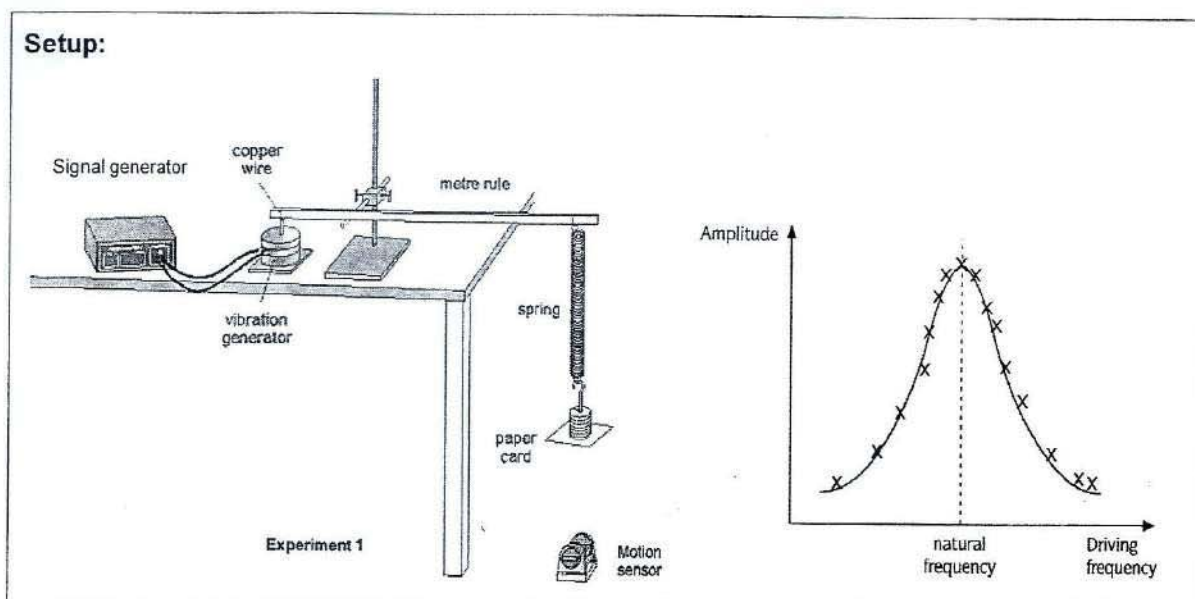
- (1) The frequency at which resonance occurs is known as the **resonant frequency**.
- (2) Under ideal conditions, i.e., if no resistive forces are present (there is no damping), the amplitude will build up to an infinitely large value.
- (3) However, in real cases, damping occurs. So, whilst putting energy into the system in each cycle, energy is also taken away by damping and the amplitude will not build to infinity.
- (4) Resonance represents the point where the (driving) energy is transferred most efficiently into the oscillatory system.



Graph (a) shows the child's x - t graph as one periodically transfers a fixed amount of energy into the swinging motion.

Graph (b) shows the variation of driving force with time.

Investigating Resonance in the Lab

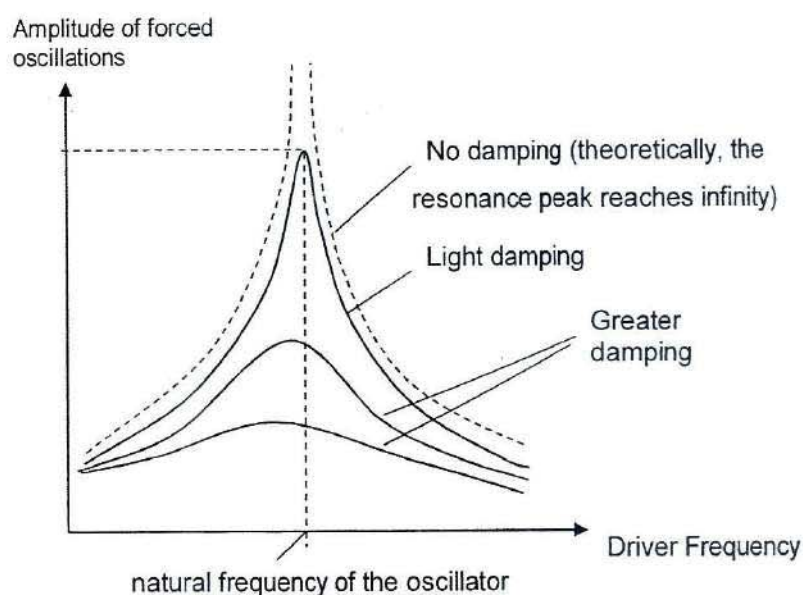


The above set-up may be used to investigate the phenomenon of resonance in the lab.

- (1) The system under study is a spring-mass system that is driven to oscillate by a metre rule which is in turn driven to vibrate by a vibration generator. To vary the frequency of vibration, the vibration generator is connected to a signal generator.
- (2) To vary the amount of damping experienced by the spring-mass system, a light paper card is attached to the mass. By varying the area of the paper card attached, the damping is varied. The larger the paper card, the greater the damping.
- (3) To measure the amplitude of vibration, a motion sensor is placed directly below the paper card.

- (4) For each frequency of vibration, we can then measure and record the amplitude of vibration of the spring-mass system. We start off with a driving frequency that is much lower than the natural frequency of the spring-mass system and slowly increase this frequency till beyond the natural frequency.
- (5) The graph above shows the typical results obtained from the experiment for one area of card.
- (6) The experiment is then repeated for different areas of cards – to produce different amounts of damping.
- (7) The data points for each area can then be plotted on the same pair of axes for comparison.

The figure below shows the amplitude vs. driving frequency graphs that will be obtained.



Note:

- 1) The amplitude of a lightly damped system is very large at resonance and the peak is sharp. The amplitude drops off rapidly when the driver frequency differs from the natural frequency of the system.
- 2) When the driver frequency approaches zero, the amplitude of the forced oscillations will be the same as the amplitude of the driver.
- 3) As the system experiences greater damping,
 - the amplitude of the oscillation **decreases**.
 - the resonance peak becomes **broad**er.
 - the resonance peak shifts slightly to the **left** to a slightly lower value of frequency.

From the graphs, it is evident that damping is a useful way to reduce the effects of resonance. Another way to avoid resonance is to ensure that the frequency of the periodic external force differs from the natural frequency of the driven system.

8.4.4 Examples of Destructive and Useful Resonance

Resonance is an important physical phenomenon that can appear in many different situations.

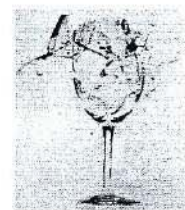
Examples of destructive resonance

1) Collapse of buildings in earthquakes

An earthquake consists of many low-frequency vibrations ranging from 1 to 10 Hz. During an earthquake, when the frequencies of the vibration match the natural frequencies of buildings, resonance may occur and result in serious damages. That also explains why some buildings collapse while others stand almost unaffected. In regions of the world where earthquakes happen regularly, buildings may be built on foundations that absorb the energy of the shock waves. In this way, the vibrations are damped and the amplitude of the oscillations cannot reach a dangerous level.

2) Shattering glass with voice

When an opera singer projects a high-pitched note whose frequency matches the natural frequency of a wine glass, the resonance may cause the glass to vibrate at so large an amplitude that it breaks.



3) Violent vibrations in vehicles and machines

If the panel in a bus rattles violently when the bus is travelling at a certain speed, it is likely that a resonant vibration is occurring. A washing machine with a load which has natural frequency matching the spinning frequency will get violent vibrations as resonance occurs.

Examples of useful resonance**1) Radio receptions**

Resonance not only occurs in mechanical oscillations but also in electric circuits. A radio receiver works on the principle of resonance. Our air is filled with radio waves of many different frequencies which the aerial picks up. The tuner can be adjusted so that the frequency of the electrical oscillations in the circuits is the same as that of the radio wave transmitted from a particular station we desire. The effect of resonance amplifies the signals contained in this wave while the radio waves of other frequencies are diminished.

2) Microwave cooking

In a microwave oven, microwaves with a frequency similar to the natural vibration frequency of water molecules are used. When food containing water molecules is placed in the oven and radiated by microwave, the water molecules resonate, absorb energy from the microwaves and get heated up. This absorbed energy then spreads through the food and cooks it. The plastic or glass containers do not heat up as much since they do not contain water molecules.

3) Magnetic Resonance Imaging

Magnetic Resonance Imaging (MRI) is increasingly used in medical diagnosis to produce images similar to those produced by X-rays. Strong, electromagnetic fields of varying radio frequencies are used to cause oscillations in atomic nuclei. When resonance occurs, energy is absorbed by the molecules. By analysing the pattern of energy absorption, a computer-generated image can be produced. The advantage of MRI scanners is that no ionising radiation (as in the process of producing X-ray images) is involved.

Appendix 1 Period of a Simple Pendulum

Restoring Force $F = -mg \sin \theta$

(-ve sign as it is in the opposite direction to the displacement)

Arc length $x = L\theta$

$$\Rightarrow \theta = x/L$$

When the angular displacement θ is small,

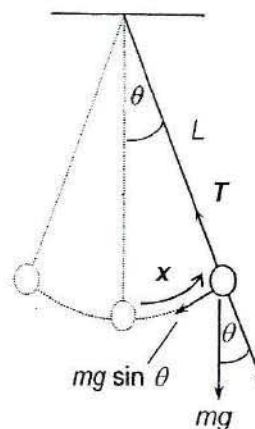
$$\sin \theta \approx \theta$$

$$\Rightarrow F = -mg \sin \theta \approx -mg \frac{x}{L}$$

From Newton's 2nd law: $ma \approx -mg \frac{x}{L}$

$$\Rightarrow a \approx -\left(\frac{g}{L}\right)x$$

i.e., $a = -\omega^2 x$ where $\omega^2 = \frac{g}{L}$



Hence, motion of a simple pendulum bob attached to a light inextensible string oscillating with a **small angular displacement of $\sim 10^\circ$ or smaller** is, to a good approximation, a linear SHM with period, T , given by

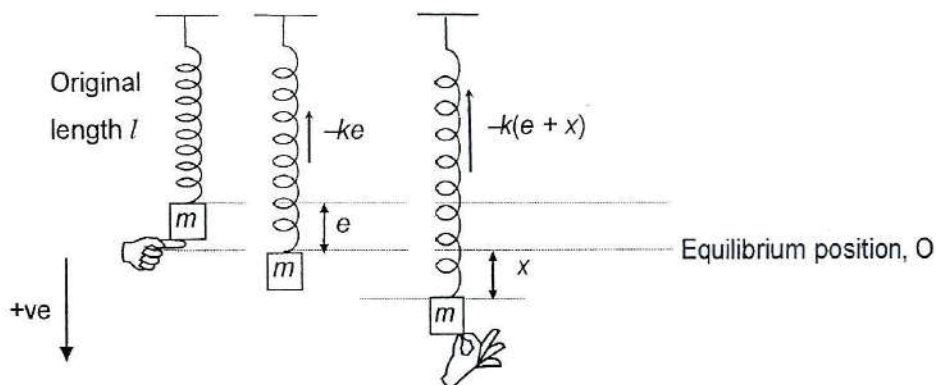
$$T = \frac{2\pi}{\omega} = 2\pi \sqrt{\frac{L}{g}}$$

Note:

- 1) Other assumptions made are that the string is massless, the radius of the bob is small compared with the length of the string and air resistance is negligible.
- 2) The period T is independent of the amplitude of the oscillation.

Appendix 2 Period of Oscillating Mass on Vertical Spring

A mass m attached to the end of a spiral spring exerts a downward force on it and stretches it by an amount e as shown in the diagram.



If the spring obeys Hooke's law, then the extension e is directly proportional to the extending tension, F_s , i.e.,

$$F_s = -ke \quad \text{where } k \text{ is a constant.}$$

Since the mass is in equilibrium, the net force on it is zero ($\Sigma F = 0$).

$$\begin{aligned} \text{Thus} \quad mg + (-ke) &= 0 \\ \Rightarrow \quad mg &= ke \end{aligned}$$

Suppose the mass is now pulled down a further distance x below its equilibrium position, then the tension in the spring is increased to $-k(e+x)$. Thus the resultant force on the mass is given by

$$\begin{aligned} F &= mg - k(e+x) \\ &= ke - k(e+x) \quad (\text{since } mg = ke) \\ &= -kx \end{aligned}$$

i.e., the resultant force is acting upwards, with the direction opposite to that of the displacement.

$$\begin{aligned} \text{By Newton's 2nd law,} \quad ma &= -kx \\ \Rightarrow \quad a &= -\frac{k}{m}x \end{aligned}$$

Since k and m are fixed, the oscillation of a mass on spring is simple harmonic.

The angular frequency ω is given by $\omega^2 = \frac{k}{m}$

$$\Rightarrow \quad T = \frac{2\pi}{\omega} = 2\pi\sqrt{\frac{m}{k}}$$

Appendix 3 Equations of motion for damped simple harmonic oscillations

In an oscillatory system, the decrease in amplitude caused by dissipative forces is called damping and the corresponding motion is called damped oscillation. The simplest case to analyse is a simple harmonic oscillator under damping, where its dissipative force is directly proportional to its *velocity*. This behaviour occurs in friction involving the flow of viscous fluid, such as in shock absorbers or sliding between oil-lubricated surfaces. The additional force on the body due to viscous drag is $f = -bv$, where b is a constant.

With a viscous drag, the net force on the body is $(-kx - bv)$.

Applying Newton's 2nd second law,

$$m \frac{d^2x}{dt^2} = -kx - bv$$

$$m \frac{d^2x}{dt^2} + b \frac{dx}{dt} + kx = 0$$

This is a second order linear differential equation with constant coefficients.

If $b < 2\sqrt{km}$, the system is under light damping and the equation of motion is

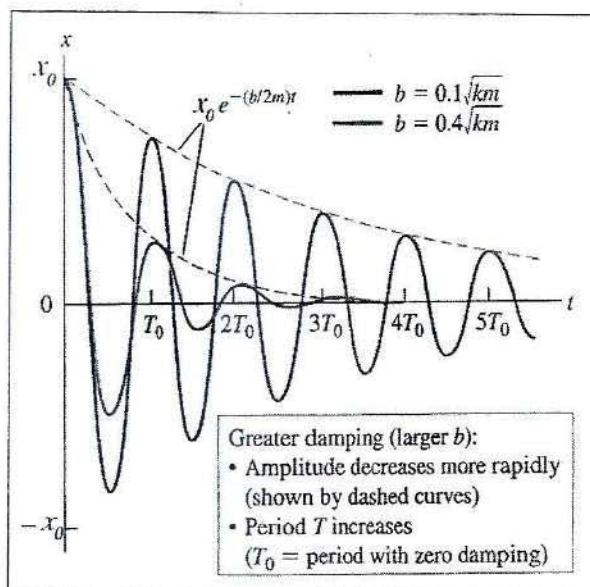
$$x = \left(x_0 e^{-\left(\frac{b}{2m}t\right)} \right) \cos \left(\left(\sqrt{\frac{k}{m} - \frac{b^2}{4m^2}} \right) t + \phi \right)$$

Its amplitude $(x_0 e^{-\left(\frac{b}{2m}t\right)})$ is not a constant, but decreases with time because of the decaying exponential factor. In addition, the larger the value of b , the faster the decrease in the amplitude.

Its angular frequency $\omega' = \sqrt{\frac{k}{m} - \frac{b^2}{4m^2}}$, is smaller than that without damping, $\omega = \sqrt{\frac{k}{m}}$, and therefore the period is slightly larger.

If $b = 2\sqrt{km}$, then $x = x_0 e^{-\left(\frac{b}{2m}t\right)}$. The system is in critical damping, where it no longer oscillates but returns to its equilibrium position without overshooting the equilibrium position.

If $b > 2\sqrt{km}$, then the system is heavily damped. Again, there is no oscillation, but the return to its equilibrium takes longer than in the case of critical damping. Its solutions are in the form of $x = C_1 e^{-a_1 t} + C_2 e^{-a_2 t}$, where C_1 and C_2 are constants determined by initial conditions and a_1 and a_2 are constants determined by m , k and b .

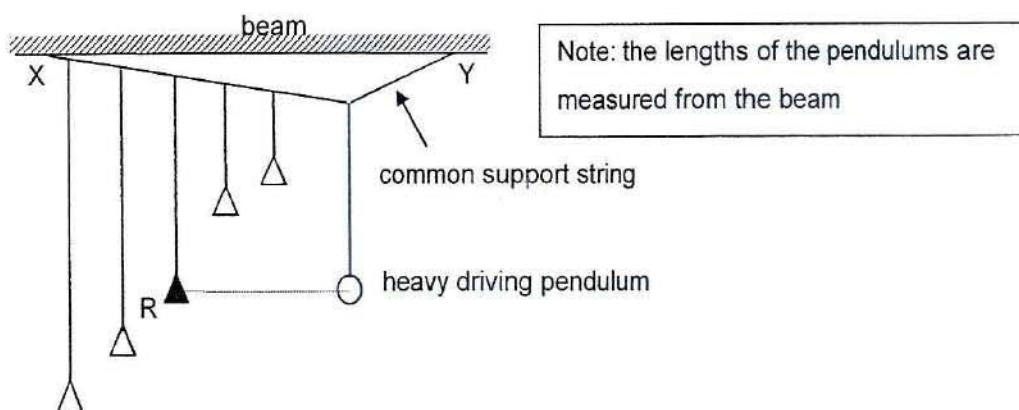


Appendix 4 Demonstration of Forced Oscillation and Resonance - Barton's pendulums

Barton's pendulums consist of several pendulums of different lengths hung from a horizontal string. Each has its own natural frequency of oscillation. The 'driver' pendulum at the end has a large mass and its length is equal to that of pendulum R. When the driver is set swinging, its vibrations are transferred through the common support string to the other pendulums and start the forced oscillations.

Observations show that:

1. All the pendulums vibrate with the same frequency as the driver.
2. Pendulum R whose length matches that of the driver pendulum (therefore has the same natural frequency as the driver) has the largest amplitude. It is said to be resonating with the driver.



Explanation:

All the pendulums are coupled together by the suspension. As the driver swings, it moves the suspension, which in turn moves the other pendulums. The matching pendulum is being pushed slightly once each oscillation, and the amplitude gradually builds up to maximum. The other pendulums are being pushed at a frequency that does not match their natural frequency and they therefore oscillate with smaller amplitudes.

Topic 9: Waves

H2 Physics C1 2016

Topic 9: Waves

Syllabus 9749

Content

- Progressive Waves
- Transverse and Longitudinal Waves
- Polarisation
- Determination of Frequency and Wavelength

Learning Outcomes

Candidates should be able to:

- (a) show an understanding and use the terms displacement, amplitude, phase difference, period, frequency, wavelength and speed.
- (b) deduce, from the definitions of speed, frequency and wavelength, the equation $v = f\lambda$.
- (c) recall and use the equation $v = f\lambda$.
- (d) show an understanding that energy is transferred due to a progressive wave.
- (e) recall and use the relationship, intensity $\propto (\text{amplitude})^2$.
- (f) show an understanding of and apply the concept that wave from a point source and travelling without loss of energy obeys an inverse square law to solve problems.
- (g) analyse and interpret graphical representations of transverse and longitudinal waves.
- (h) show an understanding that polarisation is a phenomenon associated with transverse waves.
- (i) recall and use Malus's law (intensity $\propto \cos^2\theta$) to calculate the amplitude and intensity of a plane polarised electromagnetic wave after transmission through a polarising filter.
- (j) determine the frequency of sound using a calibrated oscilloscope.
- (k) determine the wavelength of sound using stationary waves.

Content Page

Syllabus	I
Content Page	II
9.1 Introduction	1
9.2 Different Classifications of Waves	1
• Mechanical, Electromagnetic and Matter Waves.....	2
• Transverse and Longitudinal Waves.....	4
• Progressive Waves and Stationary Waves.....	5
9.3 Basic Terminology used in Waves	6
• 1D, 2-D and 3-D Waves.....	6
• Sinusoidal Waves.....	6
• Derivation of Relationship between Frequency, Speed and Wavelength:	10
9.4 Graphical Representation of Waves	12
• Graph of Displacement (y) vs. Position (x)	12
• Graph of Displacement (y) vs. Time (t)	14
9.5 Phase and Phase Difference	16
• 9.5.1 Concept of Phase and Phase Difference in Oscillation.....	16
• 9.5.2 Phase Differences in Waves:.....	17
i. Phase difference between States of Oscillation from a Displacement vs. Time Graph.....	17
ii. Phase difference between Two points on a Wave – From Displacement vs. Position Graph.....	18
iii. Phase difference between Two points of a Wave from the Displacement vs. Time Graphs of the Two points	18
iv. Phase difference between Two Waves.....	19
9.6 Energy Transmitted by a Wave	21
• Intensity of a Wave	21
• Intensity of a Wave from Point Source or Spherical Source	22
9.7 Polarisation of Waves	23
• 9.7.1 Polarisation of Light.....	24
i. Light Passing Through Two Polarisers	25
Appendices	28
• Appendix A: 3-D films and Polarization.....	28
• Appendix B: Polarisation by absorption.....	29

9.1 Introduction

So far in our study of Physics, we have been focusing very much on particle systems (i.e. a bulk concentration of mass). However, the concept of waves is another important concept in Physics.

So, what is a wave?

Consider a rope that is held under tension. If we give the left of the rope a small wiggle, the wiggle (which we call the waveform) will travel along the length of the rope. (Fig. 9.1.1)

When the waveform travels down the rope, **there is transfer of energy and momentum**, however each section of the rope only vibrates up and down about an equilibrium position, but no section of the rope actually travels from one end to the other.

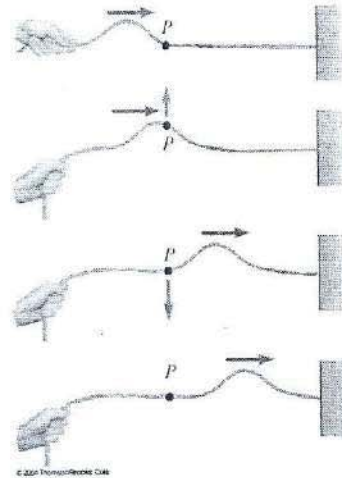


Fig. 9.1.1. Wave pulse travelling down the rope

Hence,

A wave is **disturbance (or perturbation)** that is able to **transmit energy or momentum** from a source to its surroundings. However, points in the system that carry the wave do not actually move from the source through the region the wave moves through.

9.2 Different Classifications of Waves

Waves can be categorised by three broad properties (Fig. 9.2.1):

- (i) the nature of the wave,
- (ii) mode of vibration of the points on the wave, and
- (iii) the motion of the wave.

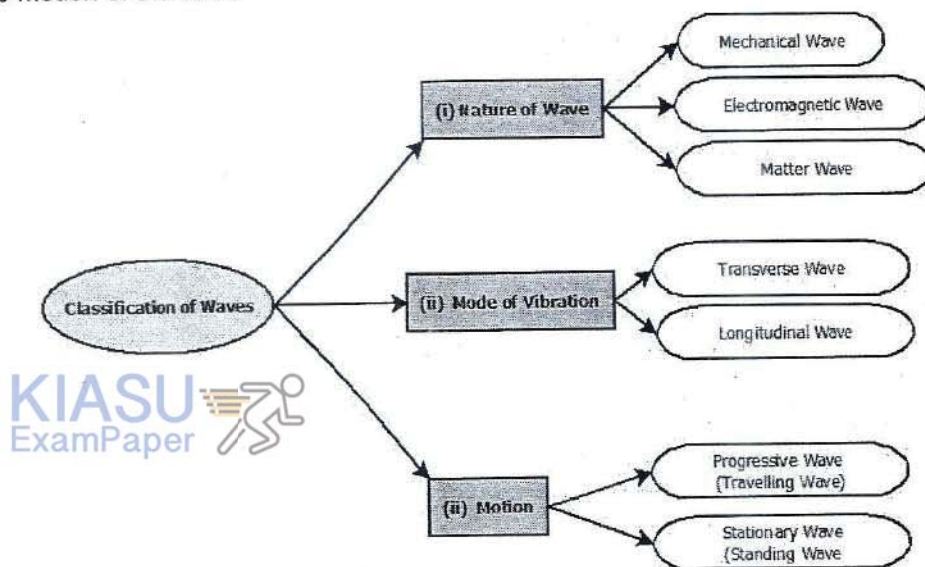


Fig. 9.2.1. Different Classifications of Waves

Mechanical, Electromagnetic and Matter Waves

• Mechanical Waves

- Sound waves, waves in ropes, water waves and seismic waves are all examples of mechanical waves.
- **All mechanical waves require a medium for propagation.**¹ The speed of wave propagation depends on the medium.

• Electromagnetic (EM) waves.

- Electromagnetic (EM) waves consist of *mutually perpendicular time-varying electric field ($\vec{E}$) and magnetic field ($\vec{B}$) oscillating perpendicularly to the direction of wave travel.* (Fig. 9.2.2)

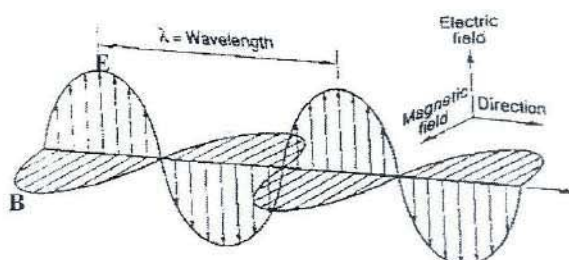


Fig. 9.2.2. The electromagnetic wave. The E-field and B-field are mutually perpendicular to each other and both fields are also perpendicular to the direction of wave travel.

- Note that both the E-field and the B-field of the EM wave vibrate at the same frequency.
- Characteristics of EM Waves:
 - All EM waves are transverse in nature, i.e. the plane of oscillation is perpendicular to the direction of motion.
 - All EM waves do not require a medium to propagate. They can move through vacuum.
 - In vacuum, all EM waves travel at the same speed c known as the speed of light:

$$c = 3.00 \times 10^8 \text{ m s}^{-1} \quad (\text{Speed of Light})$$

○ The Electromagnetic Spectrum

- Common examples of electromagnetic waves include visible light, ultraviolet light, radio wave, microwaves and x-rays. Fig. 9.2.3 shows the electromagnetic spectrum and relative positions of these EM waves on the spectrum.

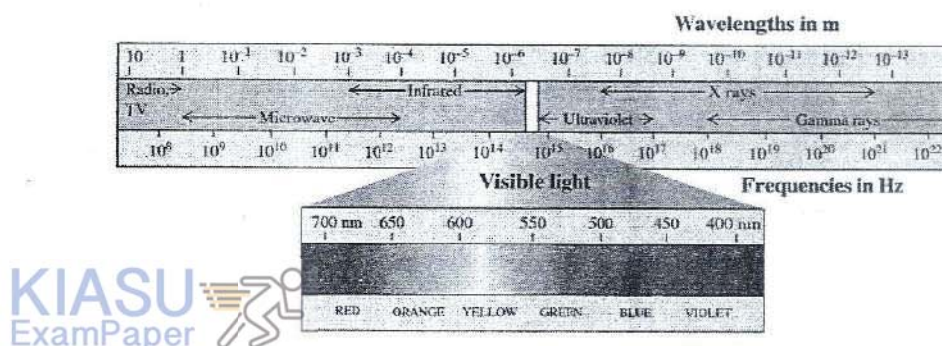


Fig. 9.2.3. The electromagnetic spectrum.

¹ Propagation of mechanical waves is possible due to the restoring forces produced upon deformation; Sound waves, which rely on oscillations of molecules of the medium resulting in pressure variation, obviously cannot be propagated through a vacuum.

○ Typical Types of EM waves

The following table shows the typical types of EM waves and their corresponding wavelengths and frequencies that you are expected to know.

Type	Wavelength (λ)	Frequency	Typical sources, uses and Effects
Gamma Rays (γ)	$\sim 10^{-15}$ m (Size of the nucleus)	$\sim 10^{20}$ Hz	From radioactive materials. Used to destroy cancer cells.
X-Rays	$\sim 10^{-10}$ m (Size of an atom)	$\sim 10^{18}$ Hz	From electrons stopped rapidly in X-ray tube. Causes ionization and fluorescence. Used for X-ray photography.
Ultraviolet (u-v) light	$\sim 10^{-8}$ m	$\sim 10^{16}$ Hz	From very hot objects, ionizes atoms. Causes fluorescence in some chemicals. Kills germs and causes suntan.
Visible light	~ 400 nm to 700 nm	$\sim 10^{14}$ Hz	Only form of radiation visible to the human eyes.
Infrared light	~ 100 μ m to 0.1 nm	$\sim 10^{12}$ Hz	From hot objects. Used for heating. Also used for very short range transmission of information (between laptops, printers, mobile phones.)
Microwaves	$\sim 10^{-2}$ m (Grids of microwave oven)	$\sim 10^{10}$ Hz	From electrons oscillating in magnetron. Used for communication (e.g. by cellular mobile phones), radar and cooking.
Radio waves	~ 1 m	$\sim 10^8$ Hz	From electrons oscillating in aerial. Used for global communication.

- A useful mnemonic to remember for visible light :

Red : Richard (700 nm)
 Orange : Of
 Yellow: York
 Green: Gained the
 Blue: Battle
 Indigo: In
 Violet: Vain (400 nm)

- The electromagnetic spectrum is a continuous spectrum i.e. there is no gaps and no frequencies anywhere in the range which do not exist. There are also actually no sharp boundaries between one type of radiation and next. For example, the wavelengths of γ -rays and x-rays overlay. Their key difference is in their mechanism of production.

- **Matter Waves** will be discussed in the topic of quantum physics.

Transverse and Longitudinal Waves

A **transverse wave** is a wave in which the points of disturbance oscillate about their equilibrium positions **perpendicular** to the direction of wave travel.

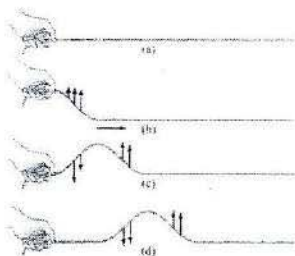


Fig. 9.2.4
Transverse Wave Pulse

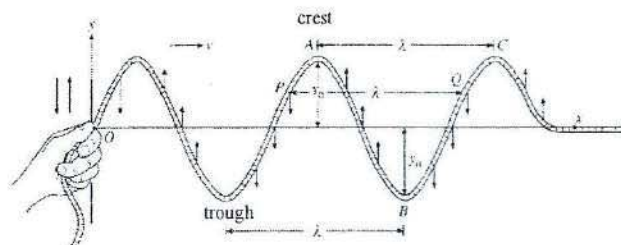


Fig. 9.2.5
Transverse Wave Train

Examples of transverse waves are waves in a rope and electromagnetic waves. Fig. 9.2.4 and Fig. 9.2.5 show a transverse waves pulse and a continuous transverse wave train on a rope respectively.

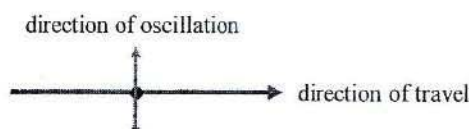


Fig. 9.2.6
Direction of oscillation of a point compared with direction of travel for a transverse wave.

A **longitudinal wave** is a wave in which the points of disturbance oscillate about their equilibrium position in a direction **parallel** to the direction of wave travel/energy propagation.

- Fig. 9.2.7 shows a longitudinal wave set up on a slinky. Sound wave is an example of a longitudinal wave.
- In a longitudinal wave, the regions in which the points or particles are closest together are known as regions of "**compression**". The regions in which the point or particles are farthest apart are known as regions of "**rarefaction**".

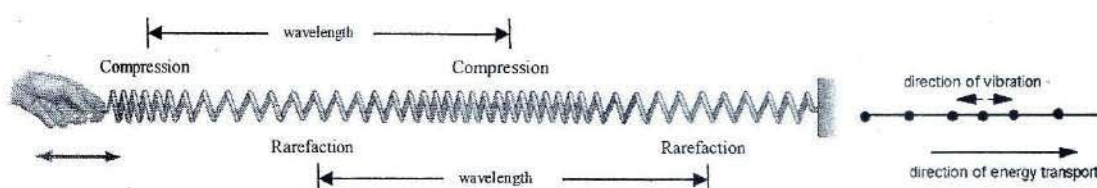


Fig. 9.2.7. A longitudinal wave train on a slinky. The direction of oscillation of a point on longitudinal is parallel to the direction of the wave travel and energy transport.

Note:

- Longitudinal waves cannot be **polarised** but transverse waves can be **polarised**.²
- Not all waves can be classified distinctively as transverse or longitudinal wave. Some waves such as seismic waves and surface water wave exhibit both transverse and longitudinal displacements.

² We will talk more about polarisation later in the chapter.

Progressive Waves and Stationary Waves

- Progressive Waves are also known as *travelling waves*.
 - Progressive waves are waves in which **the wave profile moves away from the source and causes energy to be transferred away from the source** to other regions.
 - Fig. 9.2.8 below shows a progressive transverse wave pulse that is generated in a string.

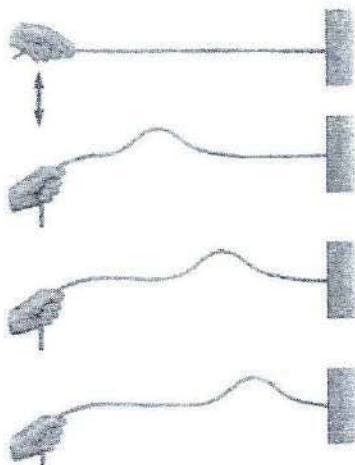


Fig. 9.2.8
Progressive wave pulse moving along a string

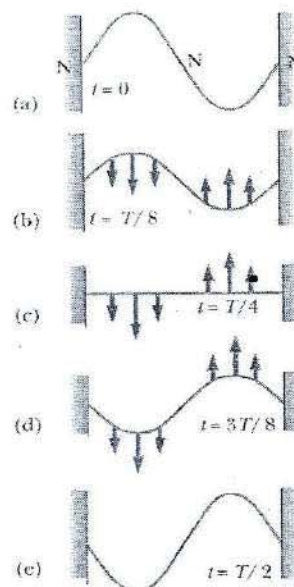


Fig. 9.2.9
A standing wave on a string

- Stationary waves** are also known as *standing waves*.³
 - A stationary wave is actually formed by the overlapping of 2 progressive waves having the same amplitude, same frequency and travelling in the same speed but moving in opposite directions in the same region of space at the same time.
 - For a stationary wave, there seems to be no progression of the wave profile in either directions (Fig. 9.2.9) and there are points on the wave (known as nodes, N) which are permanently at rest.
 - As there is no progression of the waveform, there is no transfer of energy. Energy is said to be trapped in the waves and does not propagate. The energy is said to be *localised*.

³ We will further elaborate on stationary waves and its applications in music when we have covered Principle of Superposition for waves in the next chapter.

9.3 Basic Terminology used in Waves

1-D, 2-D and 3-D Waves

- The waves that we encounter may be one-dimensional, two-dimensional or three-dimensional.
- Waves on string and slinky are examples of **one-dimensional waves**, i.e. the waves are travelling only in one dimension. A water wave that is produced by a raindrop that is striking the surface of a lake is an example of a **two-dimensional wave**. (Fig. 9.3.1) A sound wave in air is an example of a three-dimensional longitudinal wave. It consists of compressions and rarefactions of air molecules spreading out in all directions

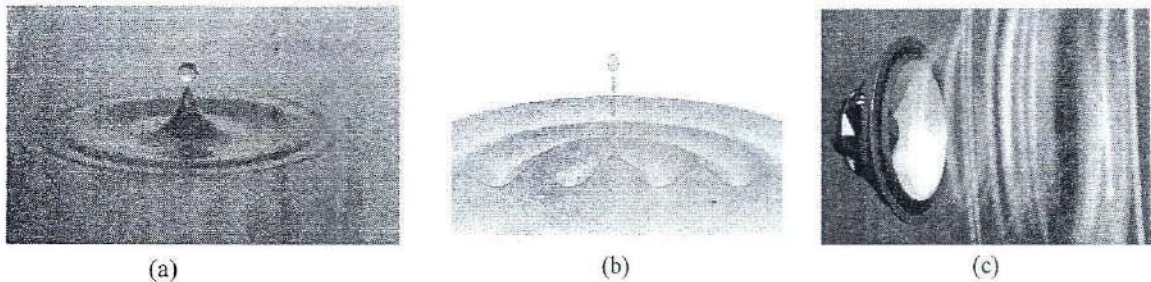


Fig. 9.3.1. (a) and (b) Two-dimensional water wave (c) Three-dimensional sound waves propagating from a speaker

Sinusoidal Waves

- Basic waveform that we will be studying is **sinusoidal waves**⁴.
- In a *sinusoidal wave*, every point on a wave is oscillating in *simple harmonic motion*.

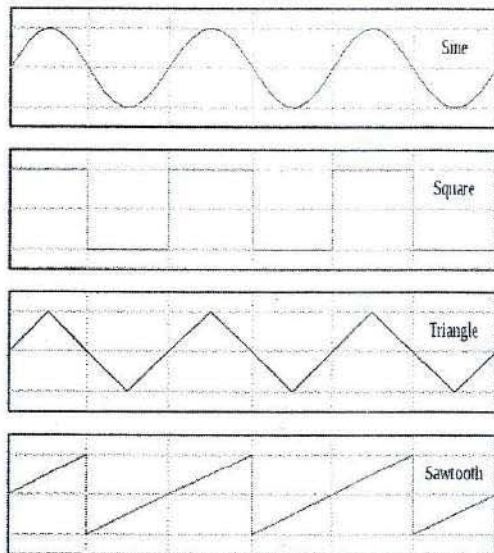


Fig. 9.3.2. Different types of Periodic Waves. All periodic waves can be formed from a combination of sinusoidal waves.

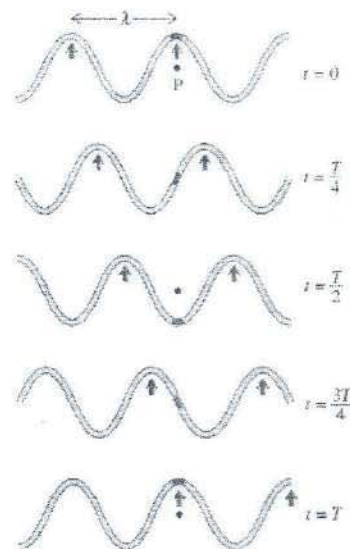


Fig. 9.3.3. As the wave progresses from left to right. (see arrow) A point (P) on a wave oscillates simple harmonically.

⁴ Any periodic wave can be represented by a combination of sinusoidal waves and hence it will be worth our attention. The technique for combining sinusoidal waves to form different waveforms is known as Fourier Transformation which you may pursue further in university.

We will use a transverse sinusoidal wave setup on the string to illustrate some basic terminology used in waves.

Suppose we set the free end of a string in simple harmonic motion. A transverse wave will be set up along the string and propagate down the string.

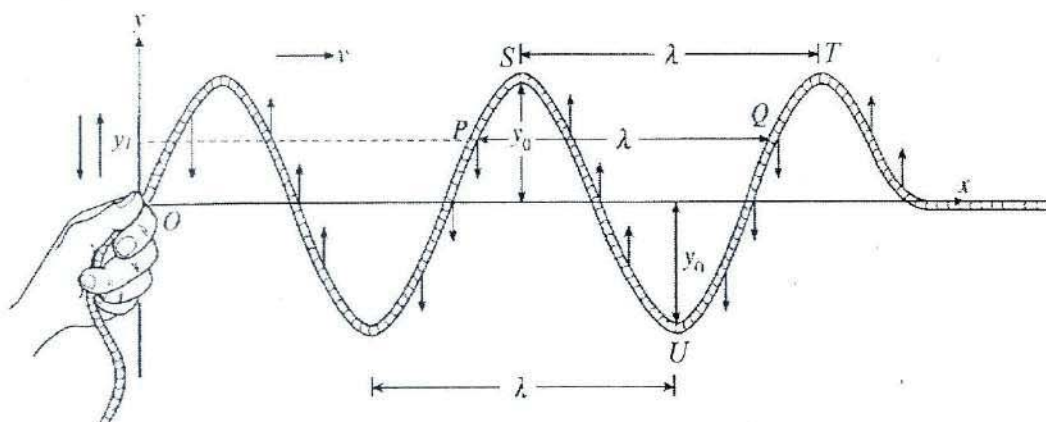


Fig. 9.3.4. A sinusoidal wave setup on rope. The figure shows a snapshot of the wave on the rope at an instant of time.

Displacement (y) :

The displacement of a particular point on the wave refers to the **distance and the direction** of the point from its equilibrium position.

Before the wave is setup, the rope lies along the x -axis. When the wave is setup in the rope, some points on the rope are being displaced from their equilibrium positions.

- Points P and Q are displaced at a distance of y_1 from their equilibrium position in the positive direction (as defined by the y -axis). Hence, their displacements are both $y = +y_1$.
- Point U is displaced at a distance of y_0 from its equilibrium position in the negative direction. Hence its displacement is $y = -y_0$.

Amplitude (A or y_0) :

The amplitude of a wave is the magnitude of the maximum displacement of a point on the wave.

- (Fig. 9.3.4). Points S, T and U are at the amplitudes of the wave.
- Note: Amplitude is always positive.

Wavelength (λ) :

The wavelength of a wave is the distance between two adjacent points which are oscillating in phase.⁵

Note:

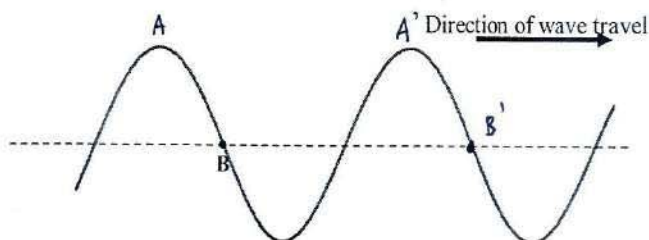
- For a transverse wave, the distance between 2 successive crests or 2 successive troughs correspond to a single wavelength.
- For a longitudinal wave, the distance between 2 successive compressions or rarefactions also corresponds to a single wavelength.

⁵ This refers to two points in the same state of oscillation. The concept of phase associated with oscillations will be dealt with further down the chapter.

Example 9.3.1:

Consider the snapshot of the wave shown below. Other than the distance between the 2 successive crests or troughs,

- identify another point (B') that pairs with B to form a single wavelength.
- Choose another pair of points on the wave that correspond to one wavelength.



- What do you notice about the pairs of points that represent one wavelength?

Same displacement, same direction of motion.

From Fig. 9.3.3, you can also see that in the time for a point on a wave to make one complete oscillation, the waveform advances by a length of a wavelength. Hence,

Period (T):

It is the time taken for a point on the wave to *complete one oscillation cycle*.

- It is also the time taken for the wave to advance by a distance of one wavelength.
- The period is also the time taken for 2 *consecutive crests/compressions or troughs (rarefactions)* to pass through a certain point.

Frequency (f):

It is the *number of oscillations per unit time* made by a point on a wave.

- It is also conveniently the no. of crests/compressions or troughs/rarefactions that pass through a point along the path of wave travel in a unit time.
- S.I. unit of frequency is given by s^{-1} or Hz (*hertz*).
- The relationship between frequency f to period T is given by : $f = \frac{1}{T}$

Wavefronts:

- The **wavefront** (of a 2-dimensional or 3-dimensional wave) is an imaginary line or curve which joins points on the wave which are oscillating in phase.
 - Therefore, *all points on a wavefront are at the same part of the cycle of their vibration.*
 - By convention, when we represent waves in a diagram, the *distance between consecutive wavefronts is one wavelength apart.*
- The direction of propagation of the waves perpendicular to the ~~direction of travel of the wave and~~ ^{wavefronts} are indicated by arrows ($\rightarrow$) known as rays.

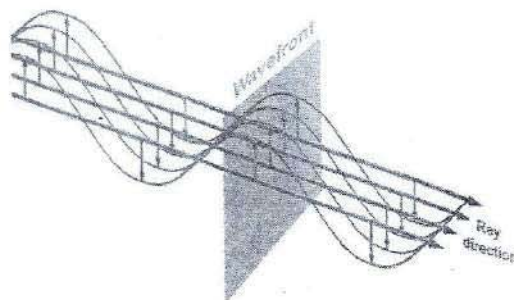


Fig. 9.3.5. Wavefront of a wave.

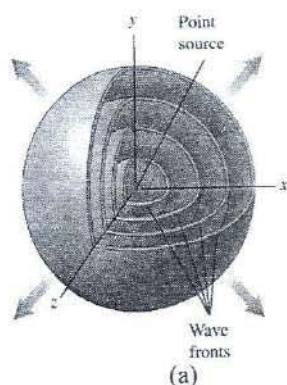
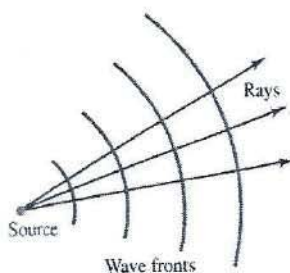
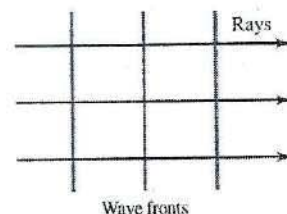
(a)
3-dimensional waves that spread out uniformly in all directions from a point source.(b)
2-dimensional representation of waves from a point source.(c)
2-dimensional representation of waves from a planar source.

Fig. 9.3.6. Wave representations using wave fronts and rays.

Wave Speed, v :

The speed of a wave is the distance that the wave profile appears to be moving per unit time.

Note:

There are actually two motions associated with waves.

- One is that of the motion associated with the oscillation of the individual points on the wave, the points always oscillate about the equilibrium position.
- The other is that of the motion associated with the propagation of waveform of the wave. This is the one that defines the wave speed.

Derivation of Relationship between Frequency, Speed and Wavelength:

The time taken for a wave to move a distance of one wavelength λ is one period T .

By definition of wave speed: Speed of the wave, $v = \frac{\text{Distance Travelled}}{\text{Time Taken}} = \frac{\lambda}{T}$

Now, $f = \frac{1}{T} \Rightarrow v = f\lambda$ where f is the frequency of the wave.

Example 9.3.2: (Sinusoidal wave)

The diagram below shows an instantaneous position of a string as a transverse wave travels along it from left to right.

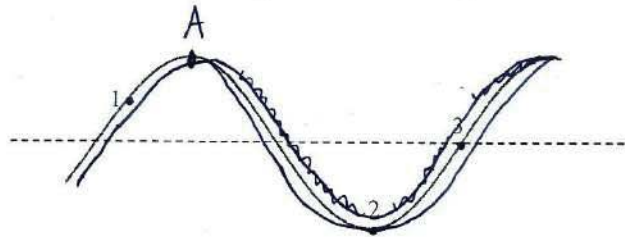


Fig. 9.3.8.

- (a) Taking upwards to be positive, for each of the points 1, 2, and 3 on the string,
(i) state whether the displacement of each point is positive or negative.

Pt 1: +ve Pt 2: -ve Pt 3: neither (zero displacement)

- (ii) What are the directions of the velocities of the points?

Pt 1: down Pt 2: instantaneously at rest Pt 3: down

- (iii) Of the three points, which one has the largest speed?

Each point is moving in SHM vertically

Pt 3 is at the equilibrium, hence largest speed

- (b) Mark on the string a point A with zero velocity.

- (c) What is the direction of acceleration of point A?

$a = -\omega^2 x$ since x is +ve, a is -ve

Example 9.3.3: Wavelength of Longitudinal Wave

Fig. 9.3.11 shows a 'snapshot' of a horizontal slinky spring when it is (a) at rest and (b) carrying a longitudinal wave moving from left to right.

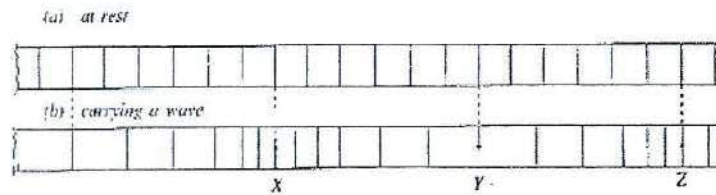


Fig. 9.3.11

State and explain whether each of the statement is true or false.

Statement A : The distance between X and Y is one wavelength.

False. XY is distance b/w centre of compression C & rarefaction R.
Hence it's 0.5λ

Statement B : The amplitude of oscillation of Y is greater than the amplitude of oscillation of X.

False. Since no energy is dissipated, amplitude of oscillation for all points is the same for a 1-D progressive wave. Hence amplitude of X = amplitude of Y.

Statement C : The displacement of X at this instant is zero.

True. X is at its equilibrium position. Hence its displacement at this instant is 0.

Example 9.3.4: Speed of a Wave

A fisherman notices that wave crests pass the bow of his anchored boat every 6.0 s. He measures the distance between the two crests to be 15 m. How fast are the waves travelling?

Answer:

Time taken for a crest to travel by 1 wavelength, $T = 6.0$ s

Distance between 2 crests, $\lambda = 15$ m

$$v = f\lambda = \frac{\lambda}{T} = \frac{15.0 \text{ m}}{6.0 \text{ s}} = 2.5 \text{ m s}^{-1}$$

Example 9.3.5

Calculate the frequency of a green light of wavelength $0.60 \mu\text{m}$.

Answer:

Speed of light, $c = 3.0 \times 10^8 \text{ m s}^{-1}$

$$\text{Using } v = f\lambda \Rightarrow f = \frac{v}{\lambda} = \frac{3.00 \times 10^8 \text{ m s}^{-1}}{0.60 \times 10^{-6} \text{ m}} = 5.0 \times 10^{14} \text{ Hz}$$

9.4 Graphical Representation of Waves

Consider a transverse water wave travelling from left to right as shown in the diagram below.

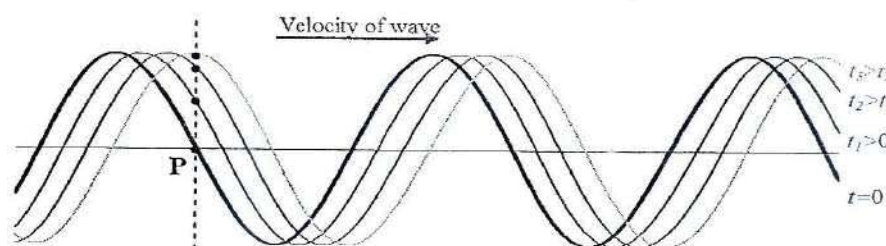


Fig. 9.4.1. A progressive transverse wave moving towards the right.

As the wave moves through this region of space, we note the following:

- At a particular instant of time, the *displacement of each point on the wave varies with the distance it is located from the source.*
- For a particular point on the wave e.g. at P, the *displacement of the point also varies with time.*

Therefore, to completely represent the wave, we need 2 graphs:

- 1 **Graph of displacement (y) vs. the position (x) of the points on a wave (at a particular instant of time.)**

Fix time $t \Rightarrow$ Graph of y vs. x .

- 2 **Graph of displacement (y) varies with time (t) for a particular point (at a particular equilibrium position).**

Fix position $x \Rightarrow$ Graph of y vs. t .

Graph of Displacement (y) vs. Position (x)

Consider a snapshot of a wave on a rope. (Fig. 9.4.2) By superimposing a pair of x and y axes on the snapshot, we can describe how the displacement of each point on the rope varies with position.

This graph of displacement vs. position at a particular instant of time (y vs. x) gives the **wave profile at an instant of time.**

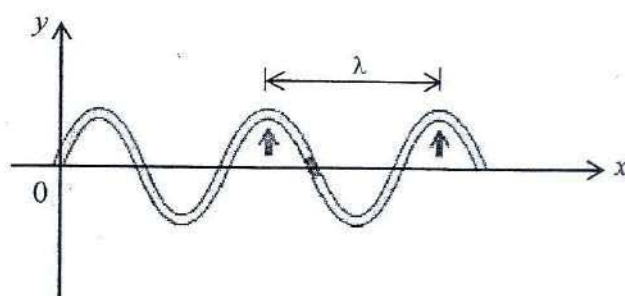


Fig. 9.4.2. Displacement-position graph of a wave on a rope.

Note :

- +ve y : displacement of a point on the wave when it is above its equilibrium position
- -ve y : displacement of a point on the wave when it is below its equilibrium position
- x axis ($y = 0$) : equilibrium position of points when there is no disturbance

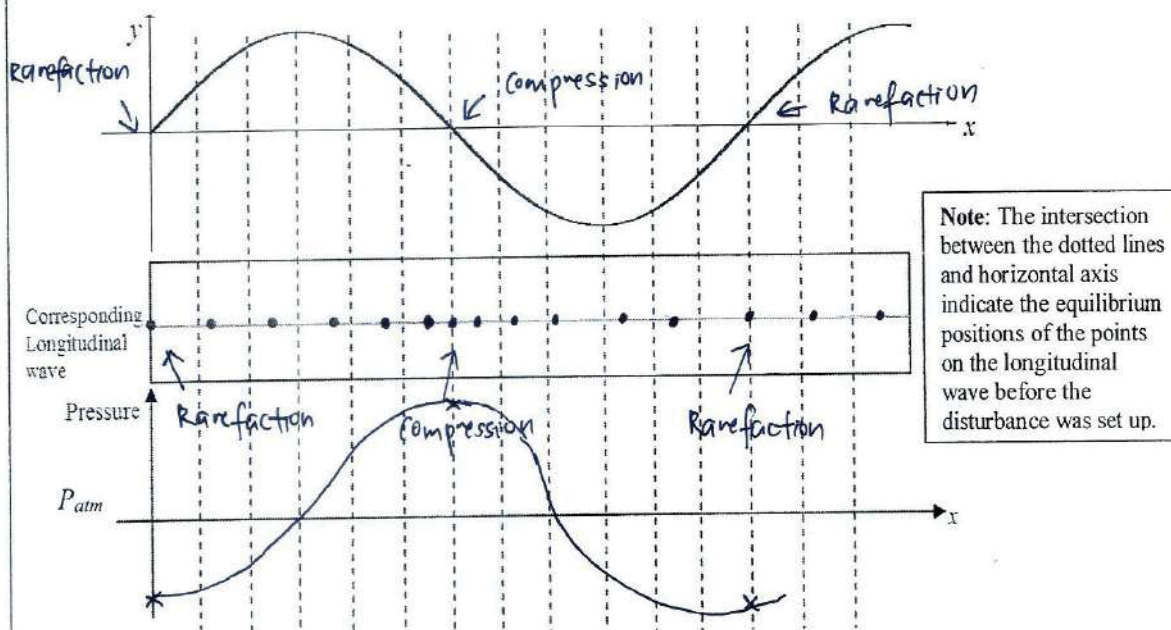
$\Rightarrow$ **Wavelength (λ) of the wave can also be obtained from the graph.**

Example 9.4.1 : Displacement-Position Graph for a Longitudinal Wave

A displacement-position graph can also be drawn for a longitudinal wave. Consider the following displacement-position graph for the air molecules along a line when a sound wave passes through a region of space.

- (i) Complete the corresponding snapshot of the longitudinal wave that is represented by the y-x graph. (i.e. indicate the positions of the air molecules). Take displacement to the right to be positive.

Sketch the corresponding pressure variation of this region.



Note: Points of compression and rarefactions corresponds to points where displacements of air molecules are zero.

Graph of Displacement (y) vs. Time (t)

Consider a point P on a rope. As a wave propagates through the rope (Fig. 9.4.3), the displacement of this point varies sinusoidally with respect to time (as shown in Fig. 9.4.4).

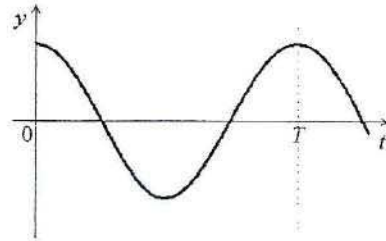


Fig. 9.4.4.

gives us

The graph of ~~displacement vs. time for a wave~~ gives us the variation of the displacement of a particular point on a wave (at a fixed position x) with respect to time.

⇒ The period (T) associated with the wave can be obtained from this graph.

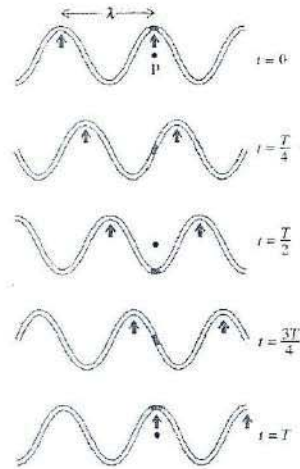


Fig. 9.4.3. A sinusoidal wave travelling on along a rope.

In summary, two graphs are needed to fully depict a wave

- The graph of **displacement vs. position** (wave profile of a wave at a certain instant of time),
⇒ Wavelength, λ , may be obtained
- From the graph of **displacement vs. time** (Motion of a particular point on a wave w.r.t. time)
⇒ Period, T , of the waves may be obtained.

Example 9.4.2: Determining speed of wave

A vibrating bar makes waves in a ripple tank. Fig. 9.4.5(a) shows the displacement of the wave as it travels out from the bar. The position of the floating cork in the tank varies with time as shown in Fig. 9.4.5(b).

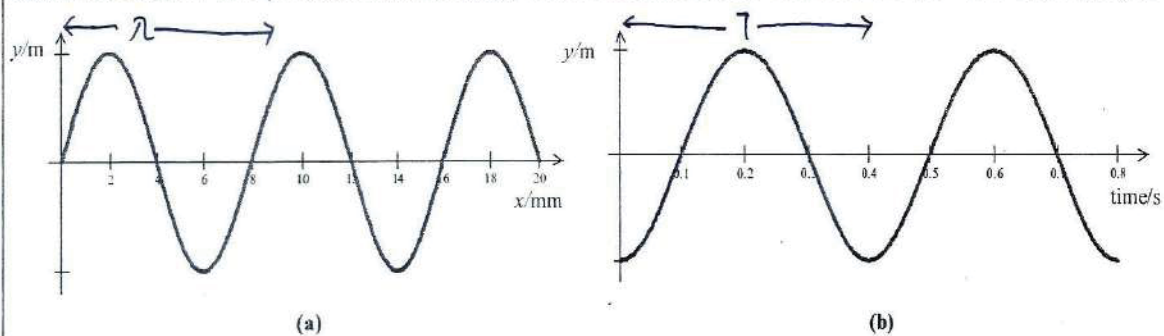


Fig. 9.4.5

From the above two graphs, determine the speed of the wave.

$$\lambda = 8 \text{ mm}$$

$$T = 0.40 \text{ s}$$

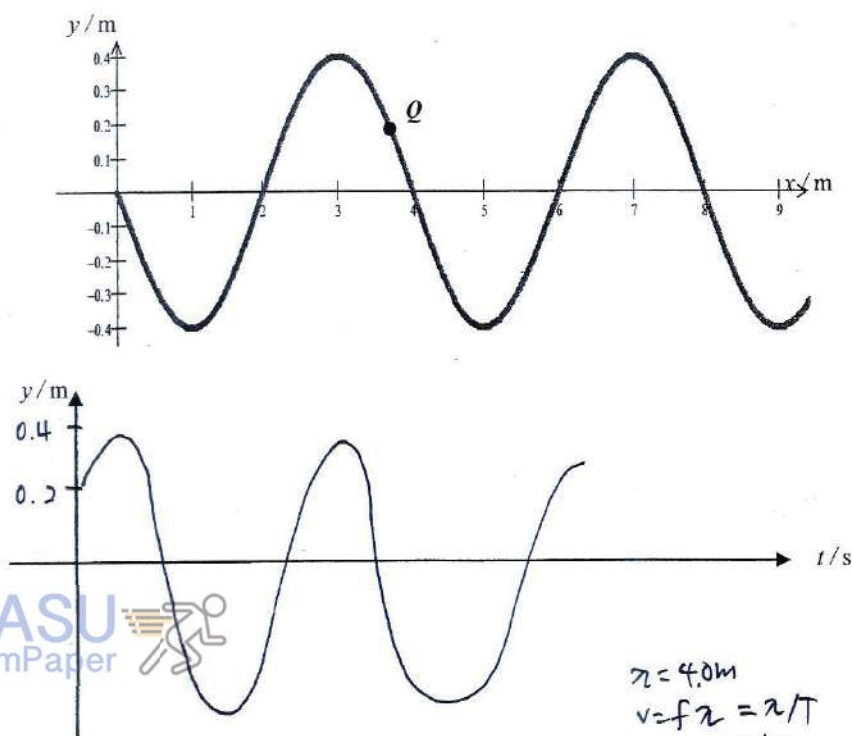
$$v = \frac{\lambda}{T}$$

$$= \frac{8}{0.4}$$

$$= 20 \text{ mm/s}$$

Example 9.4.3

The following graph shows a displacement vs. position graph of a progressive transverse wave at $t = 0$. For the point Q on the wave, sketch the corresponding displacement vs. time graph, given that the wave is moving in the positive x direction at a speed of 1.0 m s^{-1} .



$$\lambda = 4.0 \text{ m}$$

$$v = \frac{\lambda}{T} = \frac{\lambda}{T}$$

$$1.0 = \frac{4.0}{T}$$

$$T = 4.0 \text{ s}$$

9.5 Phase and Phase Difference

We have constantly referred to the term "*phase*" in both our study of simple harmonic motion and waves, however we have not formally introduced it. For this section, we will introduce *phase* by first revisiting simple harmonic motion. Then we will look at the concept of phase with respect to describing wave motion.

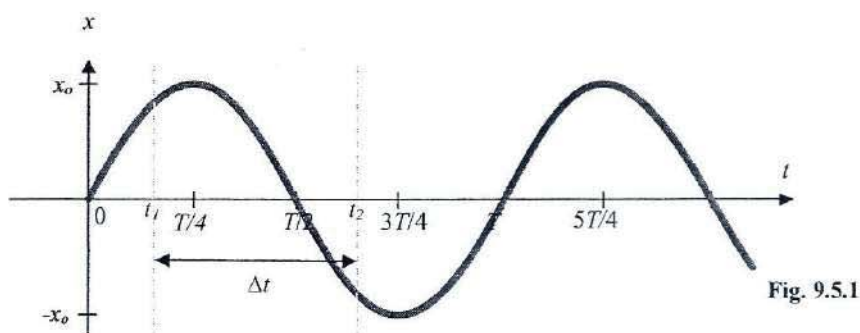
9.5.1 Concept of Phase and Phase Difference in Oscillation

When an object is oscillating simple harmonically, the displacement x of the object from its equilibrium position varies sinusoidally with time t . One of the equations that describe this displacement is:

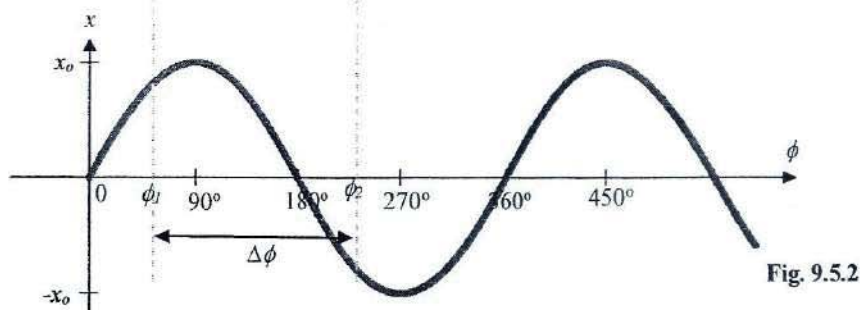
$$x = x_o \sin \omega t$$

where x_o is the amplitude of oscillation and ω is the angular frequency of the oscillation

Representing this graphically we have:



Comparing with $y = A \sin \phi$ graph:



Comparing the two graphs, we note the following

- one cycle of the oscillation (period of T) corresponds to an angle of $\phi = 360^\circ$ or 2π in radians.
- 2 points of oscillation at an arbitrary time interval of Δt apart would correspond to an angle difference $\Delta \phi$ apart.

Hence, we can further write the following:

$$T \equiv 2\pi \quad \text{and} \quad \Delta t \equiv \Delta \phi$$

$$\frac{\Delta t}{T} = \frac{\Delta \phi}{2\pi}$$

- ϕ is known as the **phase (angle)** associated with the *state of oscillation* of the particle at a particular t .
- $\Delta \phi$ is known as the **phase difference** between the 2 states of oscillation of the particle separated by a time interval of Δt .

○ **In Phase:**

When the phase difference between the 2 states of oscillation corresponds to

$$\Delta\phi = 0, 2\pi, 4\pi, 6\pi, \dots$$

the 2 states are **in phase**.

○ **In Antiphase:**

When the phase difference between the 2 states of oscillation corresponds to

$$\Delta\phi = \pi, 3\pi, 5\pi, \dots$$

the 2 states are **in anti-phase**.

In summary,

- **the phase of an oscillation** is a state of oscillation of a particle that is represented by an angle known as the **phase angle**, where 360° (2π radians) is used to represent a complete cycle, and
- **the phase difference $\Delta\phi$ between 2 points of oscillations in a time interval of Δt is given by:**

$$\frac{\Delta t}{T} = \frac{\Delta\phi}{2\pi}$$

Note that in our study, the absolute phase of an oscillation is often not important. Rather, we are interested in the phase difference between two points of comparison.

9.5.2 Phase Differences in Waves:

In general, for the phase differences in waves, we are interested in describing:

- the phase difference between two different states of oscillations for a point on a wave
- the phase difference between two points on a wave, and
- the phase difference between two waves.

Phase difference between States of Oscillation from a Displacement vs. Time Graph

For a sinusoidal progressive wave, every point on the wave oscillates simple harmonically. Hence, the phase difference between two different states of oscillation for a point on a wave that are at an interval of Δt apart from **displacement vs. time graph** is given by:

$$\frac{\Delta t}{T} = \frac{\Delta\phi}{2\pi}$$

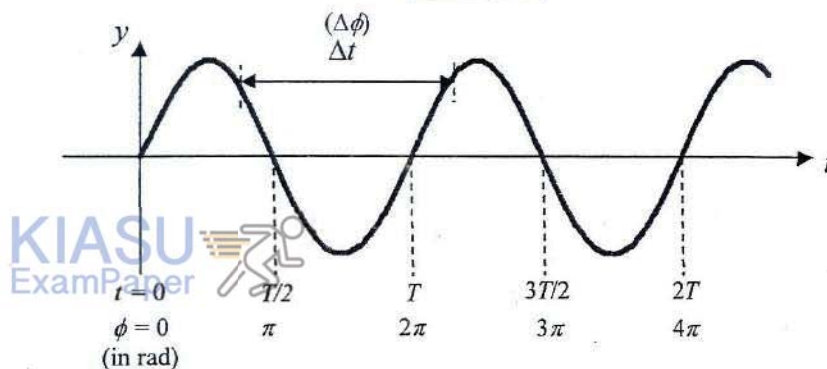


Fig. 9.5.3

Phase difference between Two points on a Wave – From Displacement vs. Position Graph

Borrowing the idea developed above, we can also represent one complete wavelength by a phase angle of 360° or 2π radians.

Therefore, the **phase difference $\Delta\phi$ between two points on a wave separated by a distance Δx** is given by:

$$\frac{\Delta x}{\lambda} = \frac{\Delta\phi}{2\pi}$$

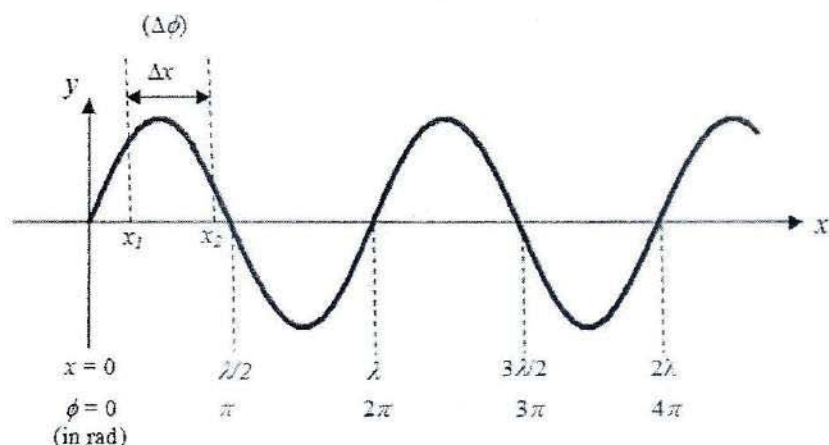


Fig. 9.5.4

Phase difference between Two points on a Wave from the Displacement vs. Time Graphs of the Two points

The displacement vs. time graph of the two points on a wave can also be used to compare the state of oscillation on the wave.

Suppose we consider the two points on the wave shown in Fig. 9.5.4 and we plot their variation of displacement vs. time graphs for the two points (one graph for each point), if the wave is progressing in the positive x direction we will get the following graphs for the points at x_1 and x_2 :

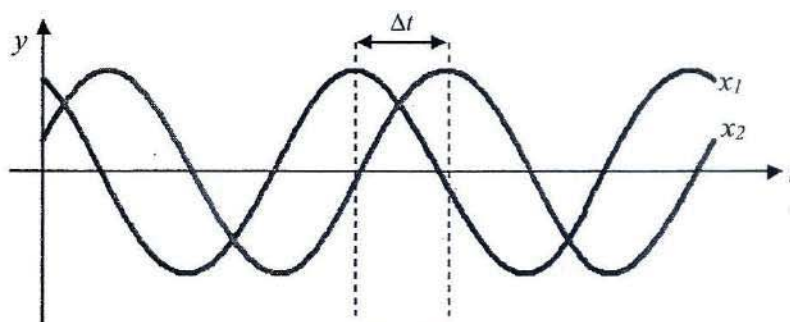


Fig. 9.5.5

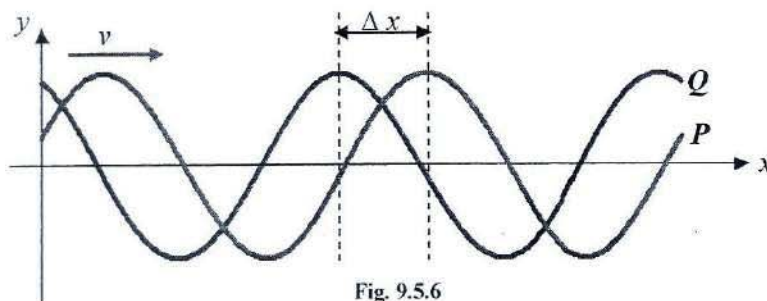
From the graphs above, we can clearly see that, the oscillation of the point at x_2 always lags behind the oscillation of the point at x_1 by a time interval of Δt , and hence point at x_2 lags behind point at x_1 by a phase angle of $\Delta\phi$ where

$$\frac{\Delta t}{T} = \frac{\Delta\phi}{2\pi}$$

Phase difference between Two Waves

We can also compare *the phase between two waves* by comparing the oscillations of every point on a wave at a particular instant.

Consider two waves *P* and *Q* having the same wavelength and velocity moving towards the right as shown in the diagram below.

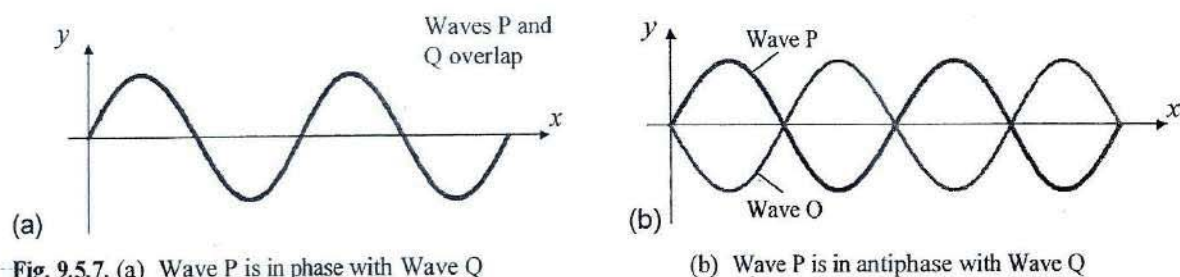


Notice that the Wave *P* always leads Wave *Q* by a constant distance of Δx . If the wave *Q* shifts a distance towards the right by Δx then every point on Wave *Q* will be in the same phase of oscillation on points on wave *P* at this moment.

Therefore, we say that

- Wave *Q* lags Wave *P* by a phase difference of $\Delta\phi$, and
- the **phase difference between the two waves** is $\Delta\phi$ can be found from : $\frac{\Delta x}{\lambda} = \frac{\Delta\phi}{2\pi}$

Note :



1. The two waves are said to be in phase if every point on Wave *P* is at the same state of oscillation as the corresponding point of the same position in Wave *Q* at the same time.

- Notice that when the 2 waves are in phase, their "*crests*" will correspond and their "*troughs*" will also correspond. The 2 waveforms overlap.

2. When the phase difference between the 2 waves are $\Delta\phi = \pi, 3\pi, 5\pi, \dots$, we say that the two waves are in antiphase.

- When the 2 waves are in antiphase, every particle in Wave *P* is in antiphase with the corresponding point.
- Note that the waveforms seem like a reflection of one another.

Example 9.5.1 :

- (a) The following graph shows the displacement vs. position of 2 waves P and Q at a particular instant of time, what is the phase difference between the two waves P and Q?

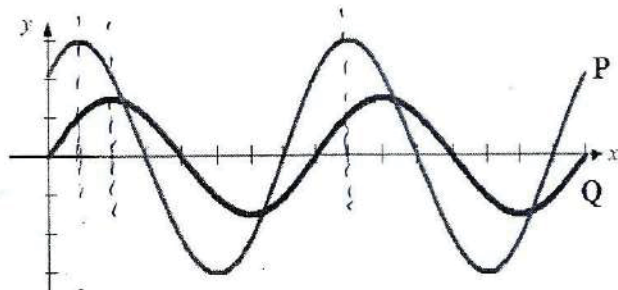


Fig. 9.5.9

$$\frac{\Delta \theta}{2\pi} = \frac{\Delta x}{\lambda}$$

$$= \frac{1}{8}$$

$$\Delta \theta = \frac{1}{8} \times 2\pi = \frac{\pi}{4} \text{ rad}$$

Q leads P by $\frac{\pi}{4}$ rad

- (b) Is it possible to compare the phase difference between two waves of different wavelengths? What is the phase difference between the two waves shown below?

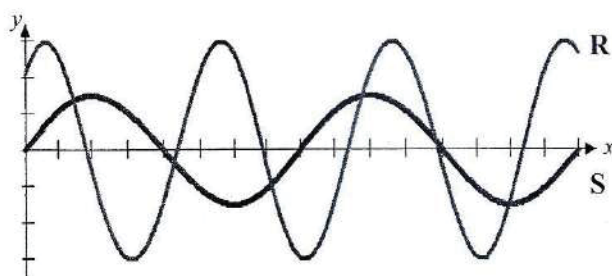


Fig. 9.5.10

Δx (dist. b/w 2 adjacent pts that are in the same state of oscillation) is not constant for the 2 waves.

The phase diff b/w the 2 waves keeps changing value (depends on position, x)

It is only meaningful to talk abt phase diff b/w 2 waves only if they have the same λ .

9.6 Energy Transmitted by a Wave

As a wave propagates, it transports energy from one place to another. Consider a sinusoidal wave moving through a string shown in Fig. 9.6.1, the source of energy is some external agent at the end of the string which does work in producing the oscillations.



Fig. 9.6.1

As the external agent performs work on the string, it transfers energy to the string in the form of kinetic energy and (elastic) potential energy to each part of the string.

From Chapter 8 Oscillations, we know that the total energy E of a simple harmonic oscillator is given by

$$E = \frac{1}{2} m \omega^2 A^2.$$

Since every part of the string is in simple harmonic motion

Energy transmitted by a wave to a point is proportional to the square of the amplitude of the oscillator at that point.

$$\text{Energy} \propto (\text{Amplitude})^2$$

Intensity of a Wave

Waves on a string carry energy in one dimension of space (along the direction of the string). But most of the waves that we commonly encounter like the sound waves in air and seismic waves in the body of the earth carry energy along all three dimensions of space. As the wave spreads out in space, the power received per unit area along the path of the wave diminishes as the distance from the source of the wave increases.

Power of a Wave:

The amount of energy transported per unit time by a wave is called the *power* of a wave.

The SI unit is joules per second (J s^{-1}) or watts (W).

Intensity:

Wave Intensity, I :

The **intensity of a wave** is the rate at which energy is transported by the wave, per unit area, across a surface perpendicular to the direction of propagation i.e. it is the power per unit area.

$$\text{Intensity, } I = \frac{\text{Power}}{\text{Area}}$$

- **S.I. unit of intensity :** W m^{-2}

- Since the energy transmitted by a wave is proportional to the square of the amplitude of the wave, so similarly, we have :

$$\text{Intensity} \propto (\text{Amplitude})^2$$

Intensity of a Wave from Point Source or Spherical Source

In the special case where the source is **point source** or **spherical source**, and the propagation is *isotropic* (i.e. the same in all directions), then the wavefronts of the waves are spherical in shape. The wave itself is said to be a *spherical wave*.

The energy spreads out over the wavefront as the wave propagates, hence, the intensity I at a particular point a distance r from the source is given by :

$$\text{Intensity, } I = \frac{\text{Power}}{\text{Area}} = \frac{P}{4\pi r^2}$$

where P is the time rate of energy delivered by the wave to a distance of r away from the source.

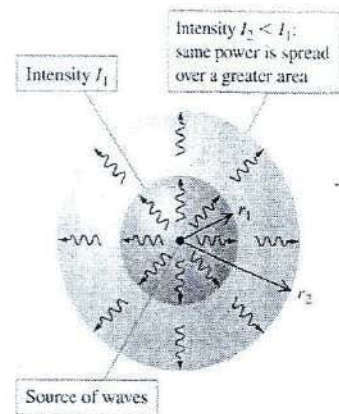


Fig. 9.6.2

Therefore, we see that the **intensity I at any distance r is therefore inversely proportional to r^2** .

$$I \propto \frac{1}{r^2} \quad (\text{For constant } P)$$

Note:

- This relationship is called the *inverse-square law for intensity*.
- This relationship is only valid for *point sources / spherical sources* and is based on the assumption that **the power of the wave does not diminish as it spreads out**.

Example 9.6.1 : (N94/II/10 – modified)

A sound wave of amplitude 0.20 mm has an intensity of 3.0 W m^{-2} .

- What will be the intensity of a sound wave of the same frequency which has an amplitude of 0.40 mm?
- How will the intensity change if the frequency of the source of wave is doubled but the amplitude remains unchanged?

Assume spherical source

$$\begin{aligned} \text{a) } \frac{I_1}{I_2} &= \frac{A_1^2}{A_2^2} \\ \frac{3.0}{I_2} &= \frac{0.20^2}{0.40^2} \\ I_2 &= 12 \text{ W m}^{-2} \end{aligned}$$

b) Source is a simple harmonic oscillator

$$\begin{aligned} \text{Since } E &\propto \omega^2 \\ E &\propto f^2 \\ I &\propto f^2 \quad (\text{Recall: } I \propto E) \\ \frac{I_1}{I_2} &= \frac{f_1^2}{f_2^2} \\ I_2 &= 4I_1 \end{aligned}$$

Example 9.6.2 :

If the intensity of an earthquake P -wave is $1.0 \times 10^6 \text{ W m}^{-2}$ at 100 km from the source, what is the intensity of the wave 400 km from the source? [Ans : $6.3 \times 10^4 \text{ W m}^{-2}$]

Assume a point source

$$I = \frac{P}{4\pi r^2}$$

Since P of source is constant $I \propto \frac{1}{r^2}$

$$\frac{I_2}{I_1} = \left(\frac{r_1}{r_2}\right)^2 = \left(\frac{100}{400}\right)^2 (1.0 \times 10^6) = 6.3 \times 10^4 \text{ W m}^{-2}$$

9.7 Polarisation of Waves

One characteristic that distinguishes *transverse waves* and *longitudinal waves* is *polarisation*⁶.

Plane Polarisation

Polarisation is a phenomenon whereby vibrations in a transverse wave are restricted to only one direction, in the plane normal to the direction of energy transfer.

To polarise a wave, we often use a *polarising filter*, or *polariser*, that permits only waves with a certain polarisation direction to pass through.

Consider waves set up on a string. A slot as shown in Fig. 9.7.1 can be used to limit the string's plane of propagation to the x - y plane and the oscillation to the y -axis. Consider waves set up on a string. A slot as shown in Fig. 9.7.1 can be used to limit the string's plane of propagation to the x - y plane and the oscillation to the y -axis.

- o **Polarising Axis or Transmission Axis**

The polarisation direction of the polariser is also known as the **polarising axis** or **transmission axis**.

Only components of the waves parallel to the polarising axis are allowed to pass through.

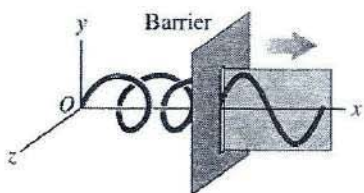


Fig. 9.7.1 A vertical slot is used to polarise the wave on the string so that oscillations are limited to a vertical plane.

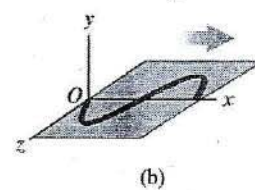
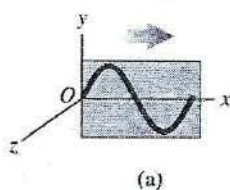


Fig. 9.7.2 A wave on a string polarised to oscillate in the x - y plane (a) and x - z plane (b). For each case, what is the polarizing axis on the polarizer?

⁶ For A-level syllabus, we will consider only plane polarisation, where we limit the oscillations to a plane. If you are interested in Circular Polarisation or Elliptical Polarisation, you may refer to Young and Freedman (11th Edition) Chapter 33.5 Pg 1268.

Why Longitudinal Waves cannot be polarised

Polarisation **does not occur** for longitudinal waves because *oscillations of the points on the wave are along the direction of wave propagation*. Restricting the direction of oscillation will in turn restrict the direction of wave propagation.

Therefore,

Longitudinal waves cannot be polarised.

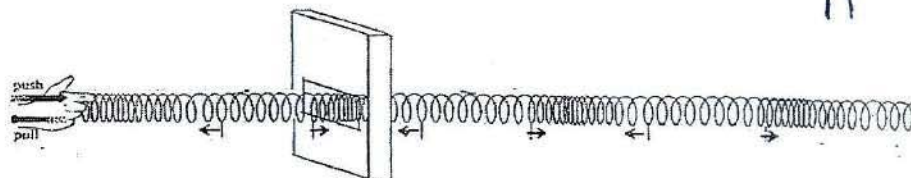


Fig. 9.7.3 Longitudinal wave on a slinky cannot be polarised as the direction of oscillation of the elements on the slinky is parallel to the direction of propagation of the longitudinal wave. Limiting the direction of oscillation will therefore also limit the motion of the wave.

In summary,

1. Only transverse waves can be polarised. Longitudinal waves cannot be polarised.
2. Only the component of the wave that is parallel to the **polarising axis** of the polariser will pass through the polarisers.

9.7.1 Polarisation of Light

Light waves are electromagnetic (EM) waves which consist of *mutually perpendicular time varying electric field ($\vec{E}$) and magnetic field ($\vec{B}$) oscillating perpendicularly to the direction of wave travel*.

EM wave is commonly represented by *just the electric field ($\vec{E}$)*. The arrows indicate the plane of oscillation of the E-field.

- Fig. 9.7.4 shows an EM wave travelling perpendicular to the plane of this paper and its plane of oscillation of the E-field is also perpendicular to the plane of the paper.
- Fig. 9.7.5 shows two ways to represent *unpolarised* EM wave which is travelling perpendicular to the plane of the paper. There are multiple planes of oscillation in an unpolarised EM wave, all of which are perpendicular to the plane of the paper as well.



Fig. 9.7.4. Representation of a polarised EM wave whose oscillation is vertical and perpendicular to the plane of the paper.

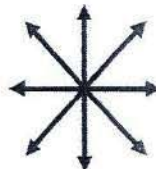
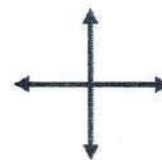


Fig. 9.7.5. Representation of an unpolarised EM wave whose planes of oscillation is perpendicular to the plane of the paper.



An ideal polariser passes 100% of the incident light that is plane-polarised along its polarisation direction and 0% of the light that is 90° to its polarisation direction. Since unpolarised light is a mix of all possible polarisations, an ideal polariser transmits only 50% of the incident unpolarised light intensity.

Light Passing Through Two Polarisers

By using two polarisers, we can also control the intensity of light.

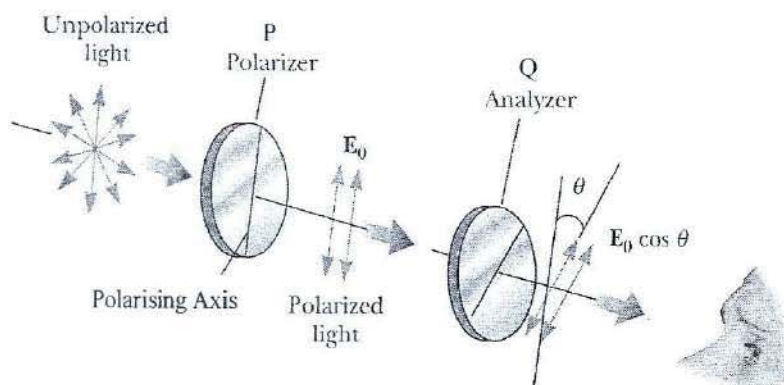


Fig. 9.7.6. Using two polarisers to control intensity of light.

Fig. 9.7.6 shows the typical setup used to control the intensity of light. An unpolarised light beam is incident on two polarisers **P** and **Q** with their polarising axes angle an angle of θ with each other.

- In such a setup, the second polariser is often known as an **analyser**.
- After passing through the first polariser **P**, the light is polarised along the polarising axis of polariser **P**. Suppose the light then has an amplitude of E_0 and an associated maximum intensity of I_0 .
- If this polarised light then falls on a second polariser **Q**, then the component of E_0 perpendicular to the analyser axis is completely absorbed, while only the component parallel to polarising axis is allowed to transmit through **Q**.
- Since the plane of oscillation of the electric field of this light is oriented at an angle of θ to the polarising axis of **Q**, therefore, amplitude of vibration of the light after passing through the polariser **Q** is given by $E_0 \cos \theta$.
- Since the intensity of a wave is directly proportional to the square of the amplitude of vibration. Therefore, we can see that

$$\frac{\text{Intensity of Light after passing through the analyser Q}}{\text{Intensity of Light after passing through the polariser P}} = \frac{I}{I_0} = \left(\frac{E_0 \cos \theta}{E_0} \right)^2$$

$$\Rightarrow I = I_0 \cos^2 \theta$$

Hence the intensity of light that passes through two polarisers that are aligned with an angle between their polarisation directions is

$$I = I_0 \cos^2 \theta$$

KIASU
ExamPaper

where I_0 is the maximum intensity of light transmitted when the two polarisers are aligned along the same direction of polarisation. The above equation is known as the **law of Malus** after its discoverer Etienne Louis Malus, who first observed this effect experimentally in 1809.

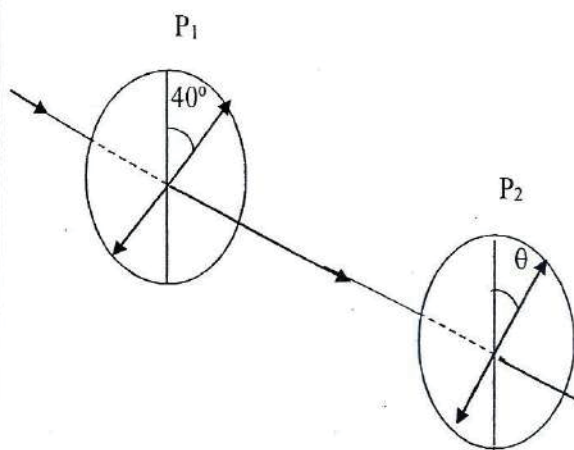
Historically polarization studies were made to investigate the nature of light. Today the procedure is reversed and we deduce something about the nature of an object from the polarization state of the light emitted by, or scattered from, that object. Polarisation studies have shown that Saturn's rings consist of ice crystals. The size and shape of virus particles can be determined by the polarization of UV light scattered from them. Much useful information about the structure of atoms and nuclei is gained from polarization studies of their emitted radiations in all parts of the EM spectrum. Thus we have a useful research technique for structures ranging in size from a galaxy ($\sim 10^{22}$ m) to a nucleus ($\sim 10^{-15}$ m). Polarised light also has many practical applications in industry and in engineering science.

Some applications of polarised light are:

1. Polarised sunglasses – Glare is usually due to light reflected off a horizontal ground (e.g. snow surface, car roof). As light gets reflected off a horizontal surface, it will be polarized horizontally. Polarised lenses contain a special filter which block off the reflected light from the ground, which reduces the glare.
2. Stress Analysis - Certain plastic materials under stress will affect the plane of polarization of light through them. This can be used to study stress in various structures.
3. 3-D films – Two images are projected through polarising filters onto a screen. The viewer will have to wear a pair of glasses which has 2 different polarising filters for each lens, so that the movie appears to be three dimensional. (see Appendix A 3-D films and Polarisation)

Example 9.7

The figure below shows a beam of initially unpolarized light passing through two polarizing filters P_1 and P_2 . The polarizing axis of P_1 is fixed at 40° with respect to the vertical axis. P_2 can be set with its polarizing axis at a variable angle θ to that of P_1 . For what value of θ do intensity minima of the emergent light occur?



For intensity minima to occur (no emergent light), the 2 polarizing axes must be $\perp$ to each other

Rotate P_2 thru 90° CW w.r.t P_1

$$\theta = 40^\circ + 90^\circ = 130^\circ$$

Rotate P_2 thru 90° ACW w.r.t P_1

$$\theta = 360^\circ - (90^\circ - 40^\circ) = 310^\circ$$

Example 9.8

Two polarising sheets have their polarizing directions parallel so that the intensity I_0 of the transmitted light is a maximum. Through what angle must either sheet be turned if the intensity is to drop by one-half?

From Malus's Law we have: $I = I_0 \cos^2 \theta$

$$\begin{aligned} \frac{I}{I_0} &= \frac{I'}{I_0} = \cos^2 \theta \\ \cos^2 \theta &= \left(\frac{1}{2} \right) \\ &= \left(\frac{1}{2} \right) \\ \theta &= 45^\circ \end{aligned}$$

Appendices

Appendix A: 3-D films and Polarization

3-D films such as Avatar (by 20th Century Fox), G-Force (by Walt Disney Pictures) and Cloudy With A Chance of Meatballs (by Sony Pictures Animation) have been a hit with the cinema these days. In fact, 3-D films have existed in some forms since 1950s. There are various techniques to produce 3-D films. Current methods include interference filter technology, for example Dolby 3D and polarization systems, for example RealD. We will look at the latter in detail.

For 3-D films which are shown using polarization systems, the viewers will have to wear 3-D glasses to view the polarized image on the screen. The image will be blurry when viewed without the special glasses. There are two types of polarization systems: linear polarization and circular polarization.

For the linear polarization system, two images are superimposed onto the same screen through two polarizing filters which are at 90 degree angles of each other (orthogonal). The silver screen is specially constructed such that it is non-depolarizing to compensate for light loss and to preserve the polarization. The viewer wears a pair of glasses which is made of a pair of orthogonal polarizing filters. Each filter only allows light from the image which is polarized in the same plane to pass through and blocks the orthogonally polarized light, each eye will only see one of the images. The 3-D effect is then achieved.



Fig. A1. Two pairs of 3-D glasses showing the polarizing effect.

Picture adapted from <http://en.wikipedia.org/wiki/File:Realdglasses.jpg>

Similarly, for a circular polarization system, two images polarized differently are superimposed onto the same screen. However, unlike linear polarization system, the two images pass through circular polarizing filter of opposite handedness. The viewer wears a pair of glasses made of a pair of circular polarizer mounted in reverse handedness. Hence, each eye will only see one of the images and the 3-D effect is achieved. RealD Cinema is a digital stereoscopic projection technology which uses similar techniques as a circular polarization system; except that it only needs one projector. A circularly polarizing liquid crystal filter which can switch polarity is placed in front of the projector lens, so that the left and right images are displayed alternately.

Information adapted from:

<http://entertainment.howstuffworks.com/digital-3d1.htm>

http://en.wikipedia.org/wiki/RealD_Cinema

http://en.wikipedia.org/wiki/3-D_film

Appendix B: Polarisation by Absorption⁷

Most light sources e.g. filament lamp, sun etc produce light that are unpolarised. However, there are ways to polarise light that is initially unpolarised, or to change the direction of polarisation of polarised light. The most common method to produce polarised light waves is by using a polariser which is known as a **polaroid**. (E.H. Land, 1928).

When light is incident on a polaroid sheet, ideally,

- if the light is polarised with its electric field **E parallel** to the polarising axis of the polaroid, it will be **completely transmitted**,
- if the light is polarised with its electric field **E perpendicular** to the polarising axis of the polaroid, it will be **completely absorbed**,
- if the light is linearly polarised at some angle θ relative to the polarising axis, the light will be partially absorbed and partially transmitted - **only the component of the electric field E parallel to the polarising axis will be transmitted**.

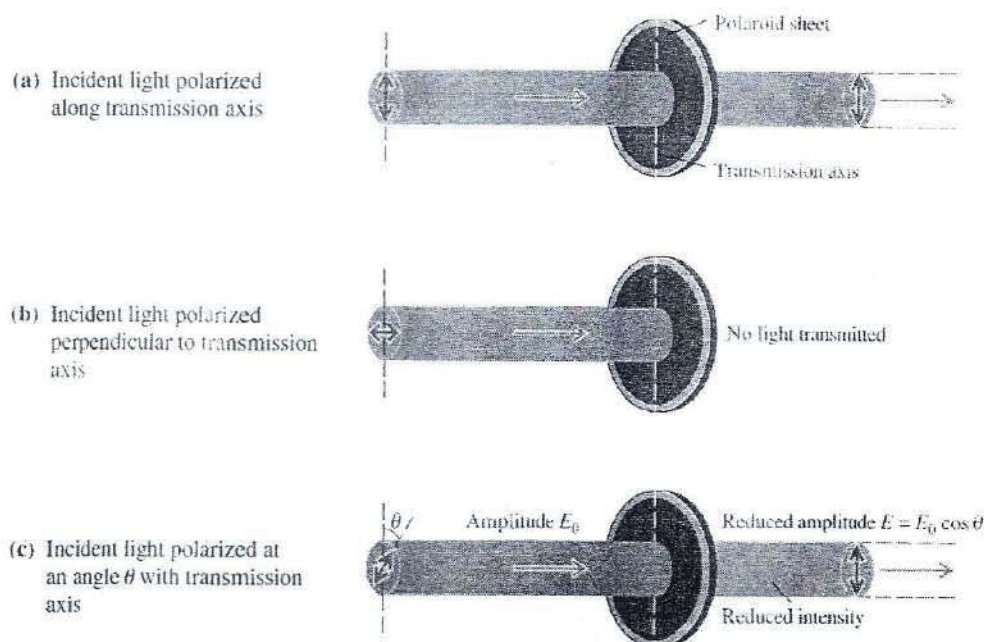


Fig. A2. The effect of a Polaroid sheet on initially polarized light depends on the direction of the initial polarisation relative to the direction of the sheet's transmission axis. (Source : Coletta, p748, Fig. 26.36.)

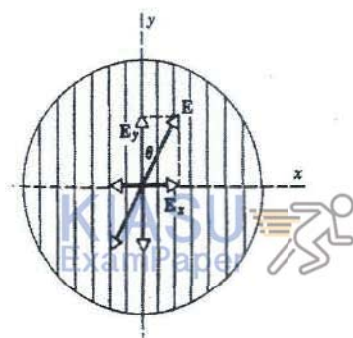


Fig. A3. The figure on the left shows what happens when polarised EM wave falls on a polariser with a vertical polarising axis. The electric field E of the EM wave makes an angle θ with the polarisation axis. The vertical component of E has magnitude $E_y = E \cos \theta$ and is transmitted. The horizontal component $E_x = E \sin \theta$ is absorbed. As a result of polarization, the amplitude of light that passes through has decreased. Hence, the intensity of the light is reduced.

⁷ It is also possible to polarise light through reflection. You may read more about it from our e-textbook "Physics Fundamentals" by Coletta, p750.

Topic 10: Superposition

H2 Physics C1 2016

Syllabus 9749

Content

- Principle of Superposition
- Interference
- Two-source interference patterns
- Diffraction
- Single slit and multiple slits diffraction
- Stationary Waves

Learning Outcomes

Candidates should be able to:

- explain and use the principle of superposition in simple applications.
- show an understanding of the terms interference, coherence, phase difference and path difference.
- show an understanding of experiments which demonstrate stationary waves using microwaves, stretched strings and air columns
- explain the formation of a stationary wave using a graphical method, and identify nodes and antinodes.
- explain the meaning of the term diffraction.
- show an understanding of experiments which demonstrate diffraction including the diffraction of water waves in a ripple tank with both a wide gap and a narrow gap.
- show an understanding of experiments which demonstrate two-source interference using water waves, sound waves, light and microwaves..
- show an understanding of the conditions required if two-source interference fringes are to be observed. :
- recall and solve problems using the equations $\lambda = a x / D$ for double slit interference using light.
- recall and use the equation $\sin \theta = \lambda / b$ to locate the position of the first minima for single slit diffraction.
- recall and use the Rayleigh criterion $\theta \approx \lambda / b$ for the resolving power of a single aperture.
- recall and solve problems by using the formula $d \sin \theta = n \lambda$ to locate the positions of the principal maxima produced by a diffraction grating.
- describe the use of a diffraction grating to determine the wavelength of light. (The structure and use of the spectrometer are not required.)

Contents

Syllabus 9749	2
Content	2
Learning Outcomes	2
10.1 Introduction	4
10.2 Principle of Superposition	4
10.3 Interference	5
10.4 Interference of Two Wave Sources	7
10.4.1 Interference of Two Wave Sources that are in Phase	7
10.5 Diffraction of Waves	12
10.5.1 Diffraction of water waves	12
10.5.2 Single Slit Diffraction Pattern	13
10.5.3 Resolving Power	16
10.6 Interference of Light Waves - Young's Double Slit Experiment	19
10.6.1 Description on Young's Double Slit Experiment	20
10.7 Diffraction Grating	26
10.8 Stationary Waves	31
10.8.1 Characteristics of Stationary Wave	32
10.8.2 Stationary Waves in Strings	34
10.8.3 Standing Waves in Air Column	36
Resonant Vibrations in an Open Pipe	38
Resonant Vibrations in a Closed Pipe	38
10.8.4 What about the pressure?	41
10.8.5 Investigating the speed of sound using stationary waves	43
Appendices	44
Appendix A: Huygens' Principle	44
Appendix B: Single Slit Diffraction Pattern	46
Appendix C: Ripple Tank	47
Appendix D: Spectrometer	48
Appendix E: Function Generator and C.R.O.	49
Function Generator	49
Cathode Ray Oscilloscope	49

10.1 Introduction

In your secondary schools, you have seen that all waves exhibit *reflection* and *refraction*.

For the A-levels, we extend our understanding of waves by looking at two other phenomena that are exhibited by waves: *interference* and *diffraction*, and also look at their applications in *Young's double slit experiment*, the *measurement of wavelengths* of light and sound, and the formation of musical notes in *string* and *woodwind* instruments through the formation of *stationary waves*.

10.2 Principle of Superposition

Imagine throwing two pebbles into a still pond. Each pebble makes its own ripples and as the ripples spread out, they kind of overlap. What determines the resultant "ripple" that we see? The displacement of the resultant wave is determined by the **Principle of Superposition**:

Principle of Superposition:

The principle of superposition states that **when two or more waves of the same kind overlap**, the **resultant displacement** at any point at any instant is given by the **vector sum of the individual displacements** that each individual wave would cause at that point at that instant.

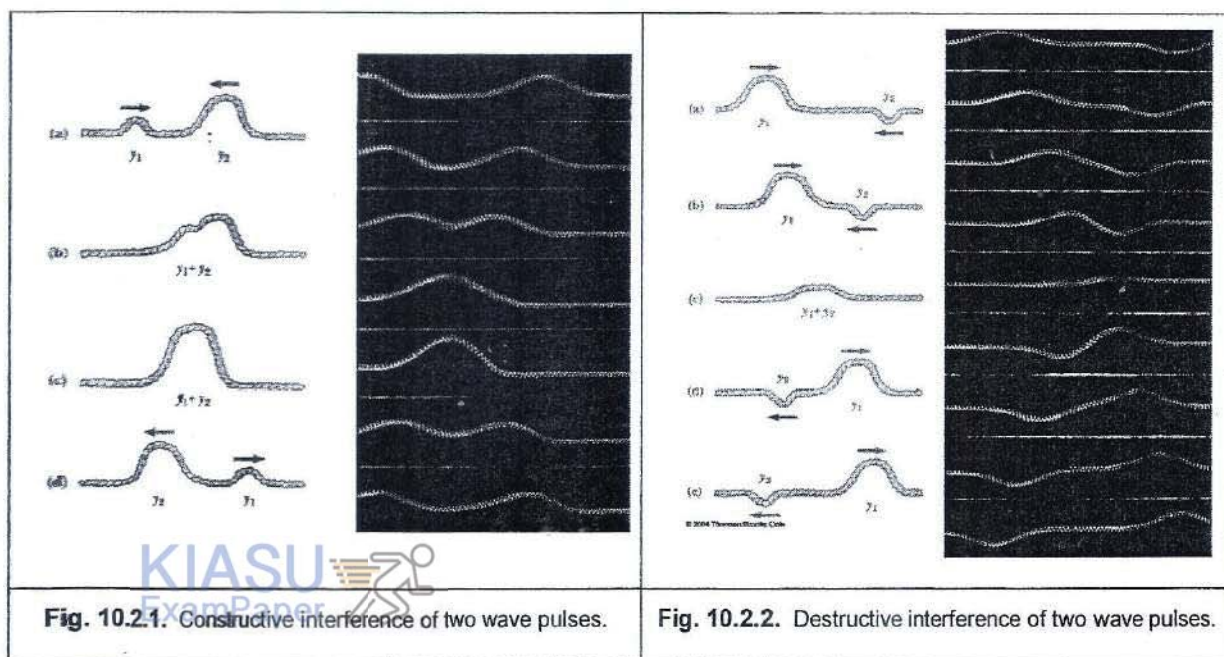
$$y_{\text{resultant}} = y_1 + y_2 + y_3 + \dots + y_N$$

where $y_{\text{resultant}}$ is the displacement of the resultant wave and $y_1, y_2, \dots, y_N$ are the individual displacements of the waves at the point of consideration.



The statement of the Principle of Superposition is a popular exam question. Be careful to use the word "**displacement**" and not "amplitude" in your statement

Fig. 10.2.1 and 10.2.2 are pictorial representations of the superposition of two wave pulses, y_1 and y_2 , travelling in opposite directions. When the waves begin to overlap, the resultant displacement of the waveform on the string at every point at any time is given by $y_1 + y_2$.



Two pulses travelling on a stretched string in opposite directions pass through each other. When the pulses overlap the net displacement of the string is equal to the sum of the displacements produced by each pulse.

Note that when the two pulses separate, they will continue to move in their original directions as if nothing has happened. From this illustration, we also see that:

After two travelling waves passed through each other, **the pulse shapes remained unchanged after the interaction, as if the two pulses had never met.**

10.3 Interference

This combination of separate waves in the same region of space at the same time to produce a resultant wave is called **interference**:

Interference

Interference is the **superposing** or overlapping of two or more waves to give a resultant wave whose **displacement** is given by the **Principle of Superposition**, which states that **displacement** of the resultant wave at any point is the **vector sum** of the **displacements** of the individual waves at that point.

Constructive interference

Constructive interference **occurs at a point** when **two waves** meet **in phase** at that point. The resultant **amplitude** of the oscillation at that point is therefore a **maximum**.

- Phase difference between the two waves at that point = $0, 2\pi, 4\pi, \dots$
- Resultant amplitude $A_R = A_1 + A_2$

Constructive Interference of Transverse Waves

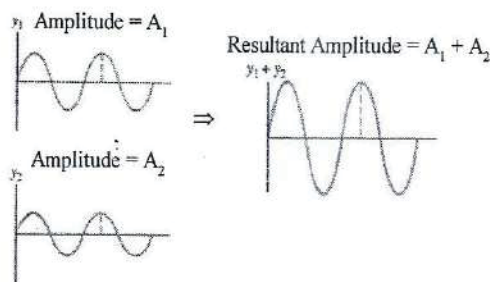


Fig. 10.3.1. Two transverse waves in phase undergo constructive interference at every point to produce a wave of maximum.

Constructive Interference of Longitudinal Waves



Fig. 10.3.2. Two longitudinal waves in phase undergo constructive interference at every point to produce a wave of increased intensity. In this case, since the amplitudes of the 2 longitudinal waves are the same, the resultant amplitude will be double that of each individual wave.

Destructive Interference:

Destructive interference **occurs at a point** when **two waves** meet **in antiphase** (phase difference of 180°). The resultant **amplitude** of the oscillation at that point is therefore a **minimum**.

- Phase difference between the two waves at that point = $\pi, 3\pi, 5\pi, \dots$
- Resultant amplitude $A_R = A_1 - A_2$
- If the waves have the **same amplitude**, then the **resultant amplitude** would be zero.

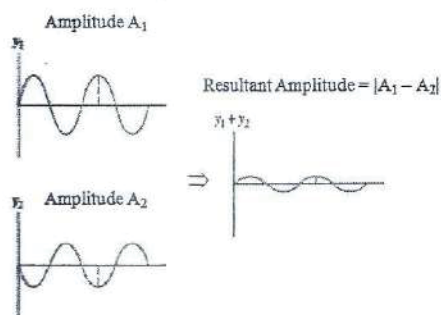
Destructive Interference of Transverse Waves

Fig. 10.3.3 Two identical transverse waves that are out of phase undergo destructive interference at every point when they are superimposed

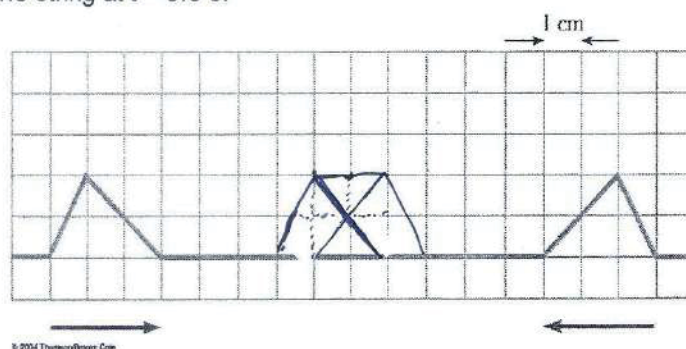
Destructive Interference of Longitudinal Waves

Fig. 10.3.4. Two identical longitudinal waves that are out of phase undergo destructive interference at every point when they are superimposed

Note that when interference is neither fully constructive nor fully destructive, it is called *intermediate interference* or *partially destructive interference*. The amplitude of the resultant waveform will be between the maximum and minimum amplitudes possible (not inclusive).

Example 10.3.1

Two pulses are travelling toward each other, at 10 cm s^{-1} on a long string, as shown in the figure below. Sketch the shape of the string at $t = 0.6 \text{ s}$.



10.4 Interference of Two Wave Sources

10.4.1 Interference of Two Wave Sources that are in Phase

Consider two identical sources of monochromatic waves S_1 and S_2 as shown in Fig. 10.4.1. The two sources produce waves of *same amplitude*, *same frequency* and are permanently oscillating *in phase*. Both waves have the same speed.

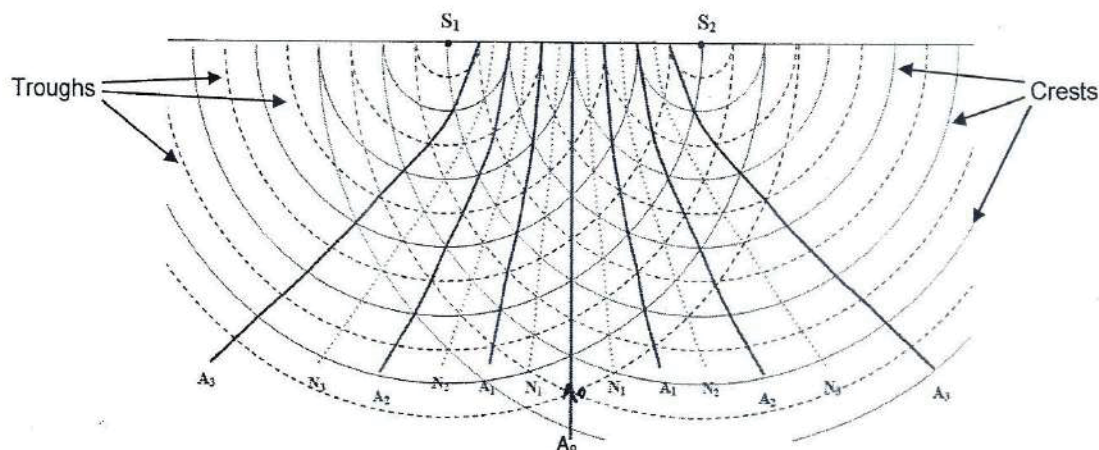


Fig. 10.4.1. Interference pattern of 2 waves sources oscillating in phase. Solid lines represent the crests while dotted lines represent the troughs of the waves produced by the two sources

Path Difference between the Two Wave Sources

Before we derive the general conditions for a *constructive interference* and *destructive interference* to be observed at any point P, we first define what we mean by *path difference*:

Path difference δ between the two waves,

$$\delta = |\text{Distance travelled by wave from } S_1 \text{ to } P - \text{Distance travelled by wave from } S_2 \text{ to } P|$$

$$= |S_1P - S_2P|$$

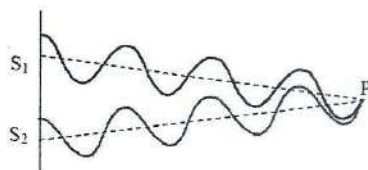


Fig. 10.4.2. Path difference between waves from S_1 and S_2 is the difference between the distances travelled by wave from S_1 and wave from S_2 .

Conditions for Constructive Interference

Consider the lines A_0 , A_1 , A_2 and A_3 .

Along these lines, the crests of the wave from S_1 (See Fig. 10.4.1) will always meet with crests from S_2 . Similarly the troughs of the wave from S_1 will meet with the troughs from S_2 .

From the Principle of Superposition, at the points where a crest meets another crest, we will get a resultant crest that is twice as high of that of the component waves. Similarly, at the points where a trough meets another trough, the resultant wave will have a trough that is twice as low as that of each component wave.

Therefore along the lines A_0 , A_1 , A_2 and A_3 , the waves will always meet in phase and hence **constructive interference** occurs. A_0 , A_1 , A_2 and A_3 are also known as the **antinodal lines**.

For a point P on the antinodal line A_0 , two waves from S_1 and S_2 travel to P over the same distance, therefore they meet in phase. We see that for all points on line A_0

$$\text{Phase difference between the 2 waves at } P: \quad \Delta\phi = 0$$

$$\text{Path difference between the 2 waves at } P: \quad \delta = |S_1P - S_2P| = 0$$

If P lies on the antinodal line A_1 , we see that the path (distance) travelled by one wave to P is always longer than the other wave by **one wavelength**.

$$\text{Phase difference between the 2 waves at } P: \quad \Delta\phi = 2\pi$$

$$\text{Path difference between the 2 waves,} \quad \delta = |S_1P - S_2P| = \lambda$$

Similarly for antinodal lines A_2 , for all points P on the line, the path travelled by one wave is always longer than the other wave by **two wavelengths**.

$$\text{Phase difference between the 2 waves at } P: \quad \Delta\phi = 4\pi$$

$$\text{Path difference between the 2 waves,} \quad \delta = |S_1P - S_2P| = 2\lambda$$

In general,

Conditions for Constructive Interference:

For two waves sources S_1 and S_2 that are oscillating **in phase**.

- **Phase difference** between the two waves meeting at point P is therefore given by

$$\Delta\phi = 0, 2\pi, 4\pi, 6\pi, \dots = 2m\pi$$

- **Path difference** between the two waves travelling from the sources to the point P :

$$\delta = |S_1P - S_2P| = 0, \lambda, 2\lambda, 3\lambda, \dots = m\lambda$$

where $m = 0, 1, 2, 3, \dots$ and λ is the wavelength of the individual waves.

Note:

1. The points of **constructive interference** are also known as points of **maxima**.
2. m is also commonly known as the **order**.
 - The central maximum is known as the **zeroth order maximum**. ($m = 0$)
 - The two maxima on either side of the zeroth order maximum are known as the **1st order maxima** ($m = 1$). The next maxima further away are then known as the **2nd order maxima** and so on.
 - Note that sometimes, we assign signs (+ or -) to distinguish the "directions" of the order e.g. the orders observed to the right of the zeroth order may be given a positive sign (+) and therefore those orders observed to left of the zeroth order may be given a negative (-) sign.

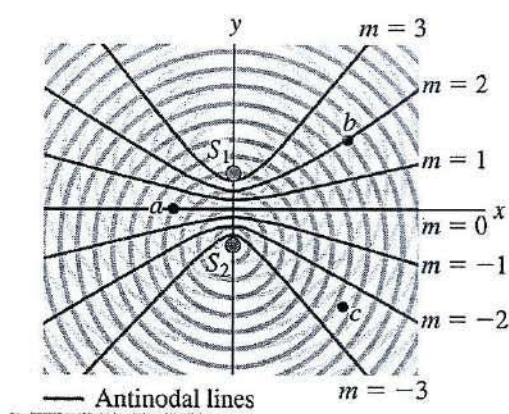


Fig. 10.4.3. Two wave sources interfere with each other to produce points of constructive and destructive interference in a two dimension area. Point *b* is a point where constructive interference may be observed. Point *c* is a point where destructive interference is observed. The antinodal lines (lines that join points of constructive interference of the same order) are indicated as well with the value of the order *m* shown. Here orders which lie above the zeroth order maximum are given a (+) sign and orders which lie below the zeroth order maximum are given a (-) sign.

Destructive Interference

Consider each point along lines N_1 , N_2 , and N_3 (See Fig 10.4.1). We notice that, along these lines, the crest of a wave from S_1 will always meet with a trough from S_2 .

Hence, the two waves will always meet in **antiphase** and therefore **destructive interference occurs**. N_0 , N_1 , N_2 and N_3 are also known as **nodal lines**.

We follow the same procedure to derive the general condition for a destructive interference to be observed at any point P from the two sources by examining the nodal lines.

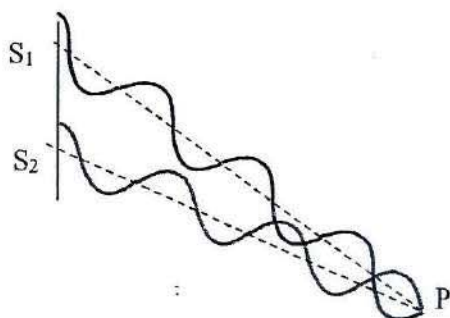


Fig. 10.4.4

For a particular *nodal line* N_1 , the wave from S_1 travels one-half of a wavelength ($\lambda/2$) more than the wave from S_2 to P , so if initially the waves were produced at the source to be in phase (phase difference is 0), this additional $\lambda/2$ will cause an equivalent phase difference of 180° and hence the two waves will meet in anti-phase. Therefore a destructive interference will be observed at P .

Now if P is a point on N_1 ,

Path difference between the 2 waves,

$$\delta = |S_1P - S_2P| = \frac{\lambda}{2}$$

If P is a point on N_2 ,

Path difference between the 2 waves,

$$\delta = |S_1P - S_2P| = \frac{3\lambda}{2}$$

Finally, if P is a point on N_3 ,

Path difference between the 2 waves,

$$\delta = |S_1P - S_2P| = \frac{5\lambda}{2}$$

Conditions for Destructive Interference:

Consider two waves sources S_1 and S_2 that are oscillating **in phase**.

For **destructive interference** to occur at a point P, a certain distance away from S_1 and S_2 ,

- **Phase difference** between the two waves meeting at point P is therefore given by:

$$\Delta\phi = \pi, 3\pi, 5\pi, 7\pi, \dots = (m - \frac{1}{2}) 2\pi$$

- **Path difference** between the two waves travelling from the sources to the point P :

$$\delta = |S_1P - S_2P| = \frac{\lambda}{2}, \frac{3\lambda}{2}, \frac{5\lambda}{2}, \dots = (m - \frac{1}{2})\lambda$$

where the order, $m = 1, 2, 3, \dots$ and λ is the wavelength of the individual waves.

Note :

1. The points of **destructive interference** are also known as points of **minima**.
2. Notice that because there is no destructive interference in the central line, there is **no zeroth order minimum observed**.

Question:

The conditions for constructive and destructive interference stated above are for the case that **the sources started out in phase**. If the original sources **are 180° out of phase** (i.e. in anti-phase or equivalent to an additional path difference of $\frac{\lambda}{2}$), will the conditions for path difference change and why?

Answer:

(The phase difference between the 2 wave trains arriving at P)

= (Phase difference at the source) + (Phase difference due to the equivalent path difference)

All the points where the waves meet in phase previously, the waves will now meet in antiphase.

All the points where the waves meet in antiphase previously, the waves will now meet in phase.

Hence, the conditions for constructive and destructive interference are interchanged.

For Constructive Interference :	For Destructive Interference :
Path difference btwn the 2 waves trains arriving at P, $\delta = S_1P - S_2P = \frac{\lambda}{2}, \frac{3\lambda}{2}, \frac{5\lambda}{2}, \dots$ $= (m - \frac{1}{2})\lambda \text{ where } m = 1, 2, 3, 4, \dots$ m: mth order maxima	Path difference btwn the 2 wave trains arriving at P $\delta = S_1P - S_2P = 0, \lambda, 2\lambda, 3\lambda, \dots$ $= m\lambda \text{ where } m = 0, 1, 2, 3, 4, \dots$ m: mth order minima

Note : In this case, the zeroth order is a minimum and there is no zeroth order maximum.

Example 10.4.1: Two Source Interference

The figure (Fig. 10.4.5) shows two sources X and Y which emit identical sound waves of wavelength 2.0 m. The two sources emit in phase, and the waves emitted have equal amplitudes, each A.

What is the amplitude of the sound wave

- (a) at R,
(b) at Q?

Suppose the source X is 180° out of phase with source Y. What does an observer hear

- (c) at R,
(d) at Q?

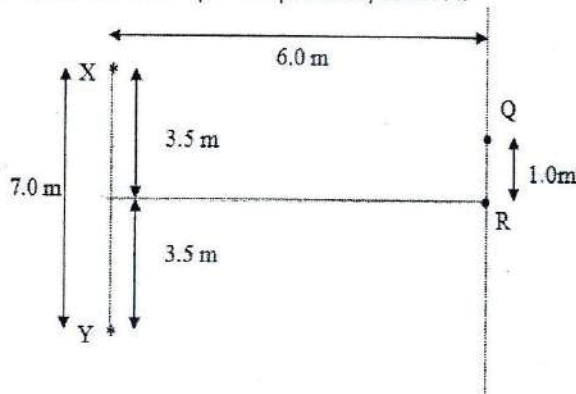


Fig. 10.4.5

(Ans: (a) 2A (b) 0 (c) 0 (d) 2A)

a) At R,

$$\text{Path difference, } \delta = XR - YR = 0$$

Therefore, constructive interference occurs

$$\Rightarrow A_{\text{resultant}} = A + A = 2A$$

b) At Q,

Using Pythagoras Theorem

$$XQ = \sqrt{6^2 + (3.5 - 1.0)^2} = 6.5 \text{ m}$$

$$YQ = \sqrt{6.0^2 + (3.5 + 1.0)^2} = 7.5 \text{ m}$$

$$\text{Path diff, } \delta = YQ - XQ = 1.0 \text{ m} = 0.5\lambda$$

$\therefore$ The waves meet in antiphase

$\Rightarrow$ destructive interference at Q

$$\Rightarrow \text{Resultant amplitude} = A - A = 0$$

c) At R,

$$\text{Path diff, } \delta = XR - YR = 0$$

$\therefore$ destructive interference occurs at R

$$\Rightarrow A_{\text{resultant}} = A - A = 0$$

$\Rightarrow$ An observer will hear no sound

d) At Q, the waves at

Q will meet in phase instead, thus

constructive interference occurs at Q.

$$\Rightarrow A_{\text{resultant}} = A + A = 2A$$

$\Rightarrow$ An observer will hear a loud sound

10.5 Diffraction of Waves

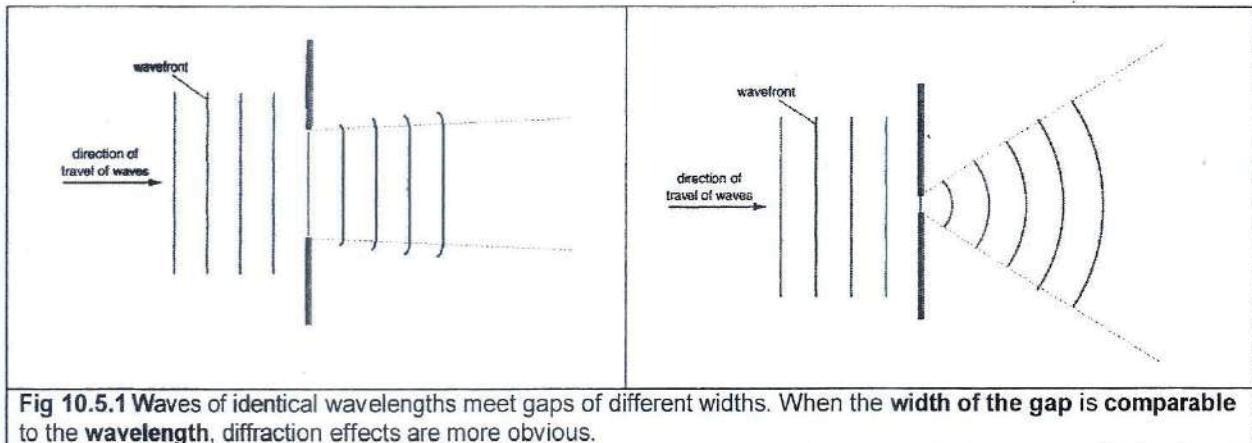
What is diffraction?

Diffraction is the **spreading** of waves, after passing through small apertures or openings, into their "geometrical" shadows; or the spreading of waves round an obstacle.

Diffraction effects become more apparent when the width of the opening is about the same size as the wavelength.

10.5.1 Diffraction of water waves

When waves meet a barrier in which there is an opening, if the width of the opening is much larger than the wavelength of the waves, the wave emerging from the opening continues to move in a straight line (apart from some small spreading at the edge). If the width of the opening is comparable or smaller than the wavelength, then the bending is a lot more obvious, and emergent wave becomes more circular.



What is meant by "Geometrical Shadow"

Consider a series of particles projected (diagram below) separately into an aperture and obstacle.



Fig. 10.5.2

As the particles are blocked by the barriers, there will be a sharp 'shadow' region where the particles do not reach. This shadow is known as the "geometrical shadow" of the barriers.

Diffraction is also observed for waves meeting an obstacle in its path. Again, **the relative size of the obstacle to the wavelength will determine the effect of diffraction** as shown in the figure 10.5.3. It is found that in order for the waves to spread around the edges of the obstacle to extend into the geometric shadow of the obstacle, the width of the obstacle should be approximately the size of the wavelength.

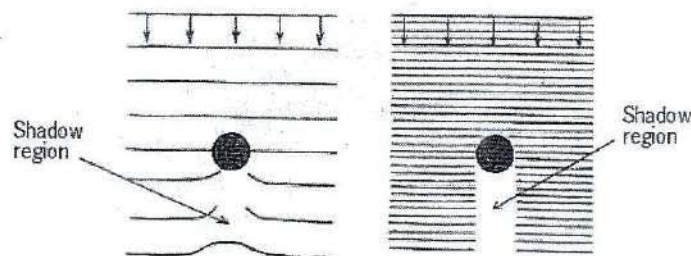


Fig. 10.5.3 The waves will bend if the length of the obstacle is smaller than the wavelength of the waves.

Since light and sound demonstrate diffraction effects. They are said to behave like waves. This diffraction effect cannot be explained by particle theory (i.e. diffraction effect is unique to waves).

10.5.2 Single Slit Diffraction Pattern

Diffraction Pattern of a Single Slit

We might expect that light passing through a small opening would simply result in a broad region of light on a screen, due to the spreading of the light as it passes through the opening. We actually find something more interesting. A diffraction pattern consisting of light and dark areas is observed, somewhat similar to the interference pattern discussed earlier.

For example, when a narrow slit is placed between a distant light source and a screen, the light produces a diffraction pattern shown in Figure 10.5.3. The pattern consists of a broad intense central band (called the central maximum), flanked by a series of alternating bright and dark bands. Figure 10.5.4 shows a diffraction pattern associated with light passing by the edge of an object. Again, we see the bright and dark fringes, similar to an interference pattern.

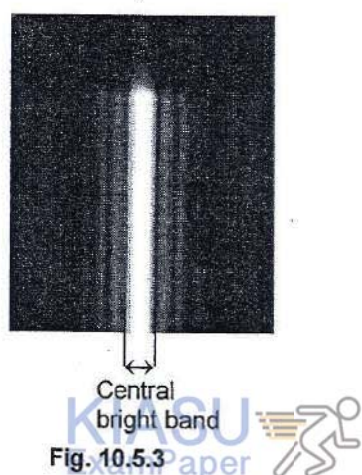


Fig. 10.5.3

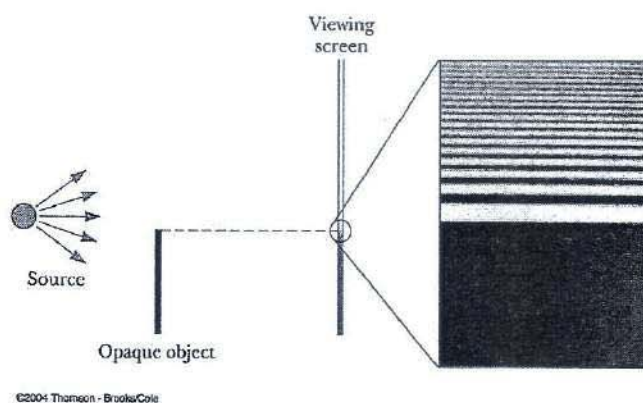


Fig. 10.5.4

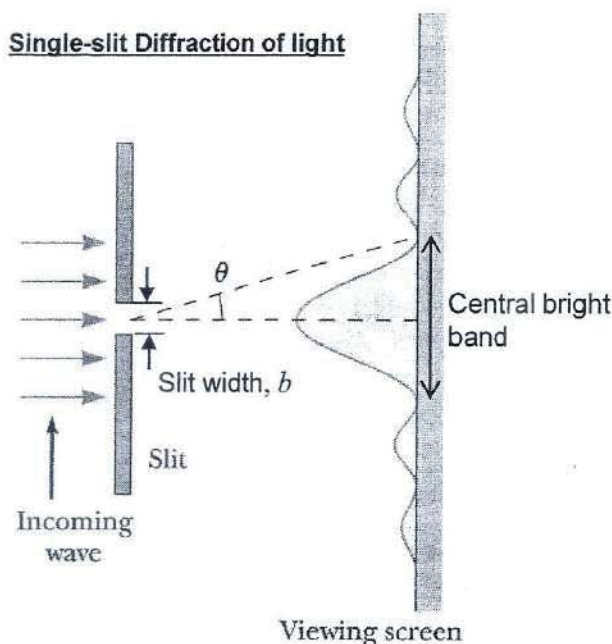


Fig. 10.5.5

In a single slit diffraction, **light waves of wavelength λ** passing through a **single slit of width b** undergoes diffraction to produce a broad intense central band with an intensity profile that is maximum at the centre, and tapers off towards both sides as shown in Figure 10.5.5.

Christiaan Huygens, a Dutch scientist, developed a geometrical model of wave propagation (known as Huygens' principle, see Appendix A) which helps in explaining the above observation. According to Huygens', the slit is made up of whole row of secondary point sources, radiating wavelets. Hence, light from one of these secondary sources can interfere with light from another secondary source. This diffraction pattern observed is actually the interference pattern produced by the many secondary light sources **within the slit**!

The **first minimum intensity** of the single slit diffraction pattern occurs at angle θ from the central maximum where

$$\sin \theta = \frac{\lambda}{b}$$

This angle θ is important because it tells us the angular spread (2θ) which contains most of the energy of the wave. It can be seen from this that **θ gets larger and the pattern spreads out as b (the slit width) is reduced**. The **pattern also gets dimmer as less light energy passes through a small slit**. A more complete analysis of the diffraction pattern can be found in Appendix B.

Width of the single slit pattern

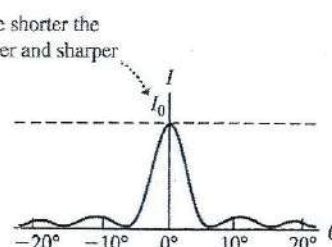
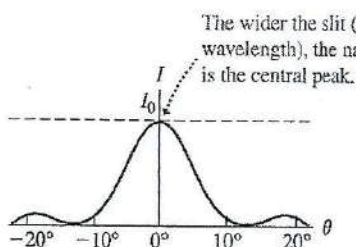
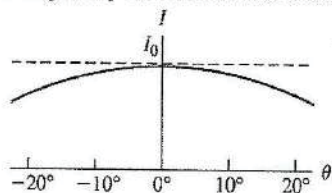
For small angles the **angular spread of the diffraction pattern** is inversely proportional to the **slit width b** or, more precisely, to the **ratio of b to the wavelength λ** . Figure 10.5.6 shows graphs of intensity I as a function of the angle for three values of the ratio b/λ .

(a) $b = \lambda$

(b) $b = 5\lambda$

(c) $b = 8\lambda$

If the slit width is equal to or narrower than the wavelength, only one broad maximum forms.

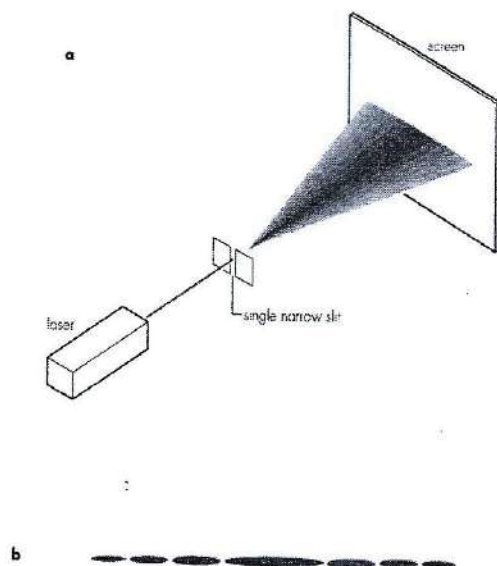


The wider the slit (or the shorter the wavelength), the narrower and sharper is the central peak.

Fig. 10.5.6 The single slit diffraction pattern depends on the ratio of the slit width b to the wavelength λ .

Example 10.5.1: Single slit diffraction

The figure (Fig. 10.5.7a) shows light from a He-Ne laser of 633 nm is incident on a 2.0×10^{-4} m wide slit and the resulting diffraction pattern on the screen as shown in Fig 10.5.7b. What is the width of the central maxima (the distance between the dark fringes on either side of the central maxima) on a screen 2.0 m away?



$$\text{Using } \sin \theta = \lambda / b$$

$$= 633 \times 10^{-9} / 2.0 \times 10^{-4}$$

$$\theta = 0.1813^\circ$$

$$\tan \theta = \frac{y}{2.0}$$

$$y = 2.0 \tan(0.1813^\circ)$$

width of the central maxima,

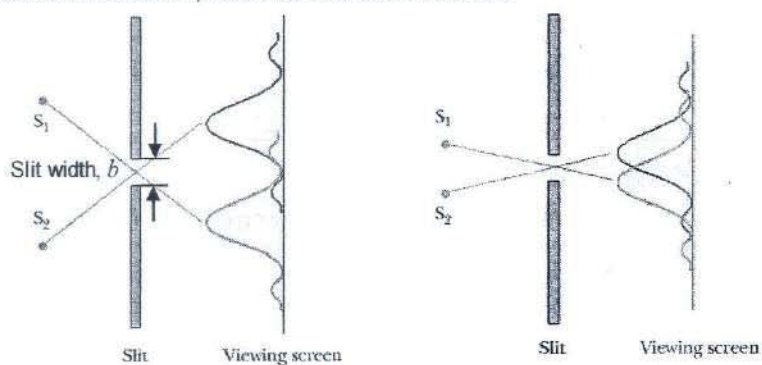
$$2y = 0.0127 \text{ m} = 12.7 \text{ mm}$$

Fig. 10.5.7

10.5.3 Resolving Power

The resolving power of an optical instrument is the ability of the instrument to differentiate clearly between the images of two close objects, i.e., the two images are seen as two distinct images. As most images are viewed through an aperture as in the case of microscopes and telescopes, it is identical to viewing the diffraction pattern produced. Hence the diffraction of light affects greatly or limits the resolving power of an optical instrument.

To understand this limitation, let us consider two light sources far away from a narrow slit of **slit width b** . The two sources can be incoherent, such as two distant stars.



(a) Little overlap of diffraction patterns

(b) Significant overlap of diffraction patterns

If no interference occurs between the two light passing through the slit, then each light source will produce a diffraction pattern on the screen. What is observed on the screen will be the summation of two diffraction patterns, one from each source. If the sources are close together, their diffraction patterns overlap; if they are close enough, their diffraction patterns overlaps almost completely and cannot be distinguished.

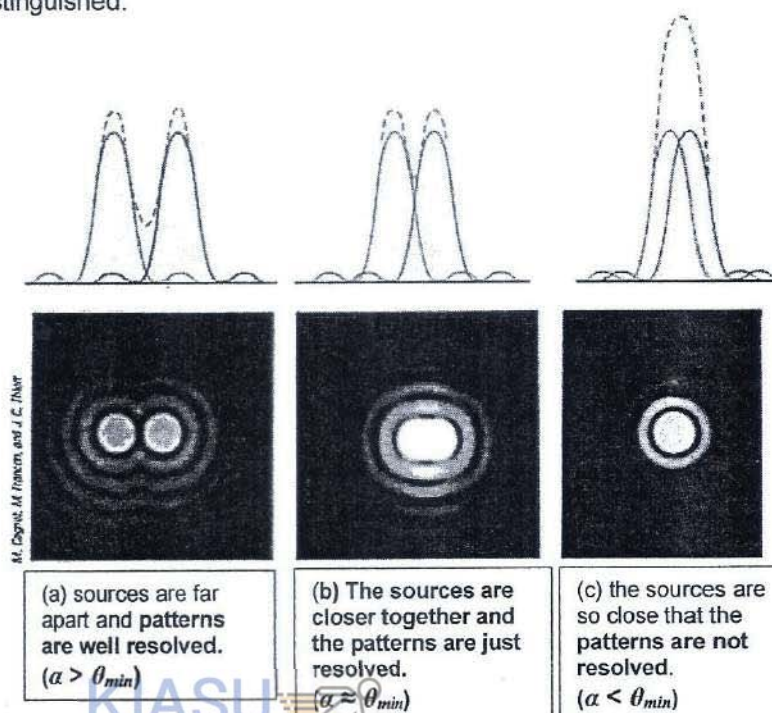


Fig 10.5.7

Individual diffraction patterns of two point sources (solid curves) and the resultant patterns (dashed curves) for various separation of sources as light passes through a circular aperture. In each case, the dashed curve is a summation of the two solid curves.

To determine if the two images are resolved, we use the **Rayleigh criterion** (named after Lord Rayleigh who suggested this rule) which states that **when the central maximum of one image falls on the first minimum of the second image, the images are said to be just resolved.**

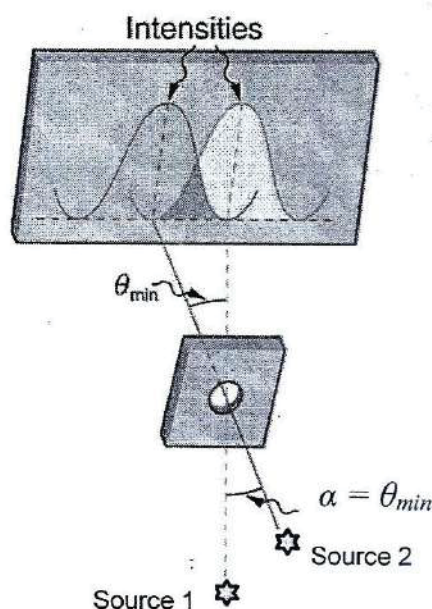
Hence according to the Rayleigh criterion, when **two images are just resolved**, the **angular separation between the two sources, α** is (approximately) equal to the **angle θ_{min}** between the central maximum and the first minimum of one of the images. (See Fig. 10.5.8) This is because the central maximum of one of the image falls on the first minimum of the second image.

From the previous section (i.e. diffraction pattern of a single slit), the **first minimum** in a single slit diffraction occurs at an angle for which $\sin \theta_{min} = \frac{\lambda}{b}$ where b is the slit width. When θ_{min} is small, we can use the approximation $\sin \theta_{min} \approx \theta_{min}$.

Therefore, the **Rayleigh criteria for the resolving power of a single aperture** is given by

$$\theta_{min} \approx \frac{\lambda}{b}$$

where b is the slit width and θ_{min} is in radians.



Hence we can also say that **the angle subtended by the two sources, α must be greater than $\frac{\lambda}{b}$ if the images are to be resolved.**

Note:

1. The resolving power of a telescope can be improved by using an objective lens of wide aperture (i.e. large diameter). Besides reducing the diffraction effects, more light is also transmitted through the telescope and a brighter image is observed.

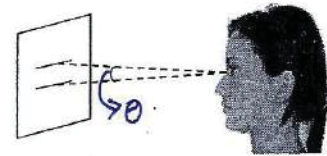
2. In microscopes, the resolving power is improved by using radiation of much smaller wavelengths than visible light, such as ultra-violet. In electron microscopes, the wavelength of the electron beam is 10^5 times smaller than visible light, resulting in much higher resolution in the images. The wavelength associated with the electron beam will be covered in Chapter 17 Quantum Physics.

Fig. 10.5.8

(Image from <http://cnx.org/contents/9ANhisjh@5/Limits-of-Resolution-The-Rayle>)

Example 10.5.3.1

The resolving power of the eye may be determined by drawing two parallel lines at a distance of say 2 mm apart on a piece of card. The card is slowly moved away from the eye until the eye just cannot see the two lines as separate lines. The distance of the card from the eye is measured. Suppose for a particular person, the distance of the card from his eye is 5 m when his eye just fail to see the two lines distinctly.



- (a) Determine the angle subtended at the eye by the two lines.
 (b) By approximating the wavelength of the visible light to be 500 nm, estimate the aperture of the pupil of the eye.

$$a) \tan \theta = \frac{2 \times 10^{-3}}{5}$$

$$\theta = 0.4 \times 10^{-4} \text{ rad}$$

$$b) \theta_{\min} = \frac{\lambda}{b}$$

$$0.4 \times 10^{-4} = \frac{500 \times 10^{-9}}{b}$$

$$b = 1.25 \times 10^{-3} \text{ m}$$

$$= 1.25 \text{ mm}$$



10.6 Interference of Light Waves - Young's Double Slit Experiment

The interference pattern produced by two *water waves* of the same wavelength can be readily seen in a ripple tank¹. Interference pattern is not easily observable when two light waves overlap, giving notion that light does not possess wave properties.

The conditions for interference to occur are:

Conditions for Interference to occur at a point:

1. Waves must be of the same kind.
2. The waves must overlap i.e. the waves must be at the same place at the same time.

So can light waves interfere with each other?

We now know that when two independent light sources are placed side by side, no interference pattern is observed because the light waves emitted by each of the sources are not coherent, i.e. they **do not maintain a constant phase relationship** with each other over the time of observation. Light from an ordinary source undergoes random phase change about once every 10^{-8} s. Therefore, conditions for constructive interference, destructive interference, or some intermediate state last for times of the order of 10^{-8} s. The result is that no interference pattern is observed since the eye cannot follow such short-term changes.

Conditions for Permanent and Observable Interference Pattern

Generally, for the interference between any 2 similar waves to be observable :

1. The sources must be **coherent**; i.e. they must maintain a *constant phase difference* with respect to each other.
2. The two wave sources must also emit waves of roughly the same amplitude.
3. For transverse waves, they must be unpolarised, or share a common direction of polarisation.

Note:

- 1 These are the conditions for **distinct (good contrast) and observable** interference pattern. If the conditions are not satisfied the waves still interfere i.e. the waves still obey the Principle of Superposition. However the interference pattern is not sufficiently pronounced or permanent enough for the observer to see.

2 Coherence:

Two waves are said to be **coherent** if they have a **constant phase difference** between them.

- When two sources are coherent, it does not necessarily imply that they are in phase with each other.
- The constant phase difference may be non-zero.
- Therefore, if 2 sources are coherent, this necessarily implies that the frequencies and wavelengths (and hence speeds) of the waves emitted must be the same.

¹ Read more about the Ripple Tank in Appendix C (Pg. 47).

10.6.1 Description on Young's Double Slit Experiment

It was Thomas Young who first demonstrated that light can also interfere with each other to produce an interference pattern as predicted by wave theory and hence demonstrating that light was a wave. A schematic diagram of the apparatus that Young used is shown in Fig. 10.6.1.

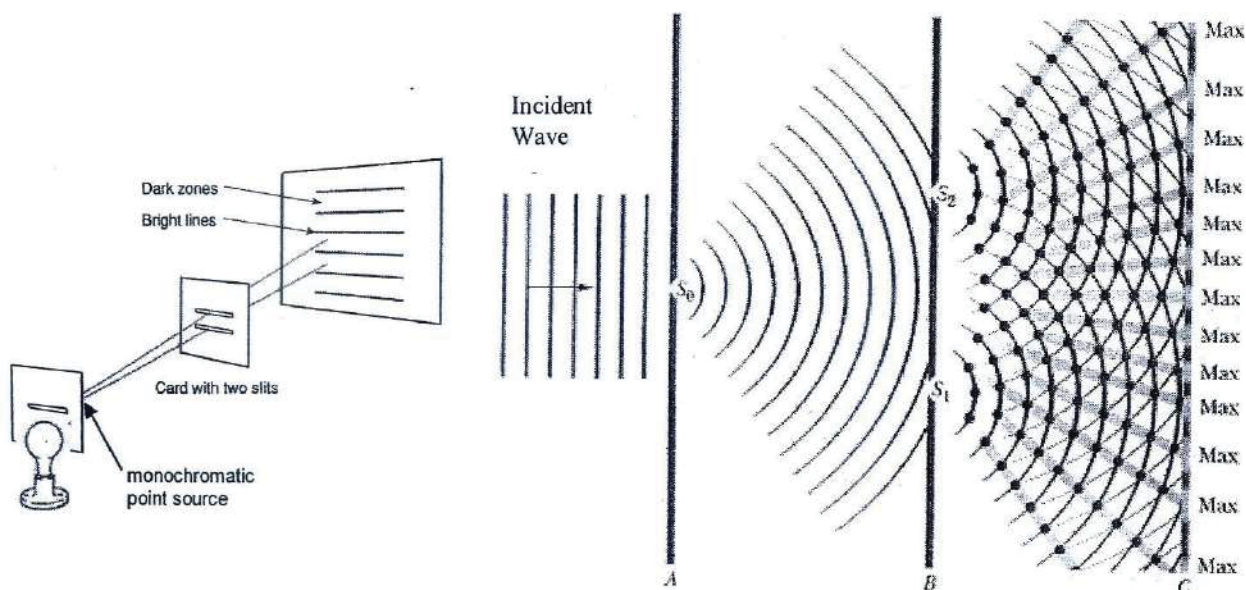
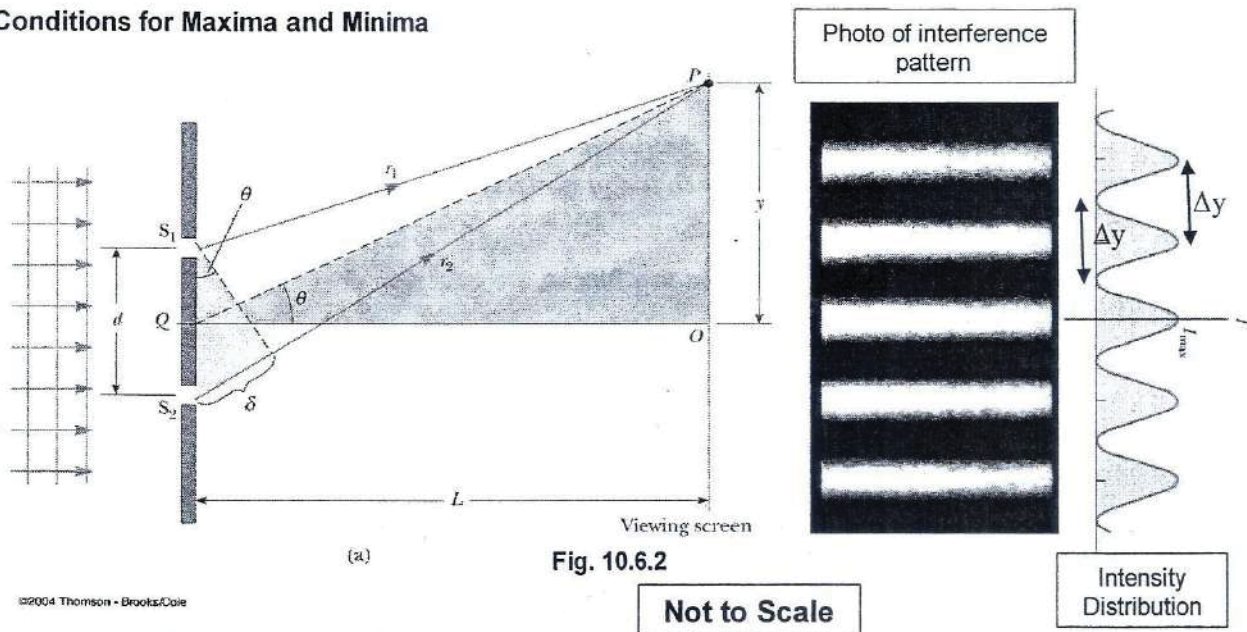


Fig. 10.6.1 Setup for Young's Double Slit Experiment

- To produce light sources which are coherent in Young's interference experiment, a point source of monochromatic light is incident on 2 slits. As light reaches the two slits, it is diffracted by slits S_1 and S_2 , which then act as 2 point sources. Since there is only one primary point source, therefore the light waves emerging from S_1 and S_2 are always coherent. If S_1 and S_2 are equidistance from S_0 , then the 2 sources are in phase.
- The light waves travelling from S_1 and S_2 overlap and undergo interference, forming an interference pattern of **maxima** and **minima**.
- On the screen points of visible bright rows called *bright bands*, *bright fringes*, or *maxima* are formed. The dark regions are called *dark bands*, *dark fringes*, or *minima*.
- The pattern of bright and dark fringes on the screen is called an **interference pattern**. (Fig.10.6.2)
- A point source is obtained usually by placing another slit S_0 between the double-slit and an extended light source.

Conditions for Maxima and Minima



Let us now derive the conditions for the positions of the maxima and minima of the Young's Double Slit experiment.

In Fig. 10.6.2 a screen is placed at a perpendicular distance L from the double-slits. The double slits are separated by a **slit separation of d** . Because the wave through S_1 originate from the same wavefront as the wave through S_2 , the two waves at S_1 and S_2 are always **in phase**.

Consider an arbitrary point P on the viewing screen,

- For a **maximum to be observed at P** , the wave from S_1 needs to **constructively interfere** with the wave from S_2 . Hence, the path difference δ between the two waves when they meet at P needs to be:

$$\delta = |r_1 - r_2| = 0, \lambda, 2\lambda, 3\lambda, \dots = m\lambda \quad \text{where } m : \text{order } (m = 0, 1, 2, 3, \dots)$$

- Similarly, for a **minimum to be observed at P** , the wave from S_1 needs to **destructively interfere** with the wave from S_2 . Hence, the path difference δ between the two waves when they meet at P needs to be:

$$\delta = |r_1 - r_2| = \frac{\lambda}{2}, \frac{3\lambda}{2}, \frac{5\lambda}{2}, \frac{7\lambda}{2} = (m - \frac{1}{2})\lambda \quad \text{where } m : \text{order } (m = 1, 2, 3, \dots)$$

However, for the experiment, the **separation between the slits d is often negligible compared to the distance L between the screen and the slits**, i.e. $L \gg d$. Therefore the rays from S_1 and S_2 to P are almost parallel to each other. (Fig. 10.6.3)

Point to Note: What are the typical orders of magnitude for slit separation, d , slit widths S_1 and S_2 and distance L for Young's double slits experiment with light?

Therefore, the **path difference** δ may be approximated by:

$$\delta = d \sin \theta$$

where d is the **separation between the slits**, and θ is the **angular displacement of the position of interference fringe observed from the principal axis QO**.²

The conditions that defines whether **bright** or **dark** fringes are observed at point P is thus given by

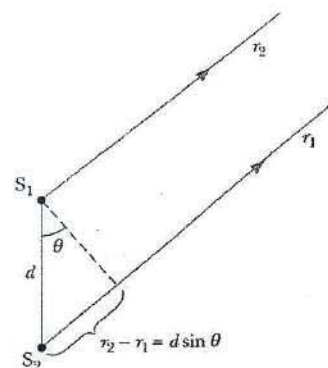


Fig. 10.6.3

Condition for Constructive Interference (Maxima or Bright Fringes):

$$d \sin \theta = m\lambda \quad \text{where } m : \text{order number } (m = 0, 1, 2, 3, \dots)$$

Condition for Destructive Interference (Minima or Dark Fringes):

$$d \sin \theta = \left(m - \frac{1}{2}\right)\lambda \quad \text{where } m : \text{order number } (m = 1, 2, 3, \dots)$$

The assumption made during the derivation of these two equations is $L \gg d$

Note:

- ii Similar to the two source interference introduced earlier, m is called the **order number**.
 - o $m = 0$ corresponds to $\theta = 0$. In this case, a central bright fringe is observed and hence it is known as the **zeroth order maximum**. The first maximum on either side is called the **first-order maximum**, and so forth.
 - o Sometimes a sign is assigned to the order to help distinguish the positions of the maxima or minima. For example in this case, maxima located at one side (suppose above) of the zeroth order maximum may have values of $m = 1, 2, 3, 4, \dots$ while maxima located at the other side (below) the zeroth order maximum may have values of $m = -1, -2, -3, -4, \dots$

2 Positions y of Bright and Dark Fringes

It is often useful to obtain expressions for the positions y along the screen of the fringes measured vertically from O to P.

In a typical Young's Double Slit experiment, constructive interference occurs when

$$d \sin \theta = m\lambda \Rightarrow \sin \theta = m\lambda/d$$

Given small values of m , the value of θ would be very small if $\lambda \ll d$,

hence we can use small angle approximation ($\sin \theta \approx \tan \theta$) to arrive at $\tan \theta = m\lambda/d$

From triangle OPQ (Fig. 10.6.2), we have $\tan \theta = \frac{y}{L}$

$$\Rightarrow \frac{m\lambda}{d} = \frac{y}{L}$$

² The **principal axis** is a line joining the midpoint of the two slits perpendicularly to the viewing screen.

Hence,

Positions of Bright Fringes: $y_{\text{bright}} = m \frac{\lambda L}{d} \quad (m = 0, \pm 1, \pm 2, \pm 3, \dots)$

Positions of Dark Fringes: $y_{\text{dark}} = (m - \frac{1}{2}) \frac{\lambda L}{d} \quad (m = \pm 1, \pm 2, \pm 3, \dots)$

The assumptions made during derivation of the above 2 equations are:

- θ is small. (Applies for low values of m where $m\lambda \ll d$)
- $L \gg d$

3 Fringe width or Fringe Separation, Δy :

The **fringe width or fringe separation** is defined to be the *distance between 2 consecutive bright fringes* or the *distance between 2 consecutive dark fringes*.

Hence, **Fringe separation**

$$\begin{aligned} \Delta y &= \text{Position of } (m+1)^{\text{th}} \text{ order bright fringe} - \text{Position of } m^{\text{th}} \text{ order bright fringe} \\ &= y_{m+1} - y_m \\ &= (m+1) \frac{\lambda L}{d} - m \frac{\lambda L}{d} \end{aligned}$$

i.e.

$$\Delta y = \frac{\lambda L}{d}$$

where λ is wavelength of light;

L is the distance of screen from slits;

d is the slit separation.

Notice that on the right hand side of the equation, λ , L and d are constants, hence we can make the following conclusion:

For a two-source interference, where $d \ll L$ and $\lambda \ll d$, the fringe separations close to the central maximum (θ is small) is approximately **constant.**

This conclusion is further confirmed by the photo in Fig. 10.6.2, which captures the fringe pattern formed on the viewing screen during an experiment. The corresponding plot of the **intensity distribution** is also given.

Example 10.6.1 : Determination of wavelength of light using Young's Double Slit

A screen is separated from the double-slit source by 1.2 m. The distance between the two slits is 0.03 cm. The second-order bright fringe ($m = 2$) is measured to be 4.5 mm from the centre line

- (i) Determine the wavelength of light.
 (ii) What is the fringe separation between 2 neighbouring bright fringes formed on the screen?

(i). Using $d \frac{y}{L} = m \lambda$

$$\lambda = \frac{dy}{mL} = \frac{(4.5 \times 10^{-3})(0.03 \times 10^{-2})}{2(1.2)} = 5.63 \times 10^{-7} \text{ m}$$

(ii). Fringe separation,

$$\Delta y = \frac{\lambda L}{d} = \frac{(5.625 \times 10^{-7})(1.2)}{(0.03 \times 10^{-2})} = 2.25 \times 10^{-3} \text{ m}$$

Effects of Diffraction on Double Slit

The slits used in the double slit experiment in practice cannot be considered as point sources. As a result, each slit produces a diffraction pattern as shown in Figure 10.5.5. Figure 10.6.4 illustrates the interference pattern of two slits superposed on the intensity pattern of a single slit. The interference fringe pattern due to the double slit gets moderated by the diffraction pattern.

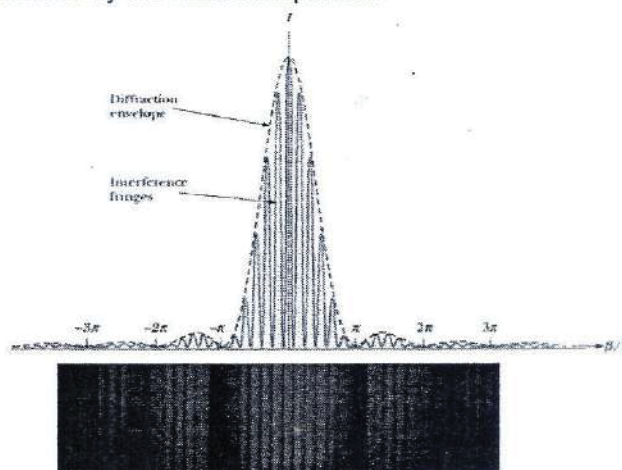


Fig. 10.6.4

This can sometimes result in missing orders in the interference fringe pattern. When the minima of the diffraction envelope coincide with the maxima of the interference pattern, the "maxima" of the interference pattern go missing.

10.7 Diffraction Grating

We have seen that the Young's double-slit experiment is able to allow us to determine the wavelength of light. However, a more commonly used tool is the **diffraction grating**. A diffraction grating has many fine, parallel, closely spaced and equidistant slits/rulings inscribed on a sheet of glass or metal. A typical diffraction grating has several hundreds to thousands of slits per mm.

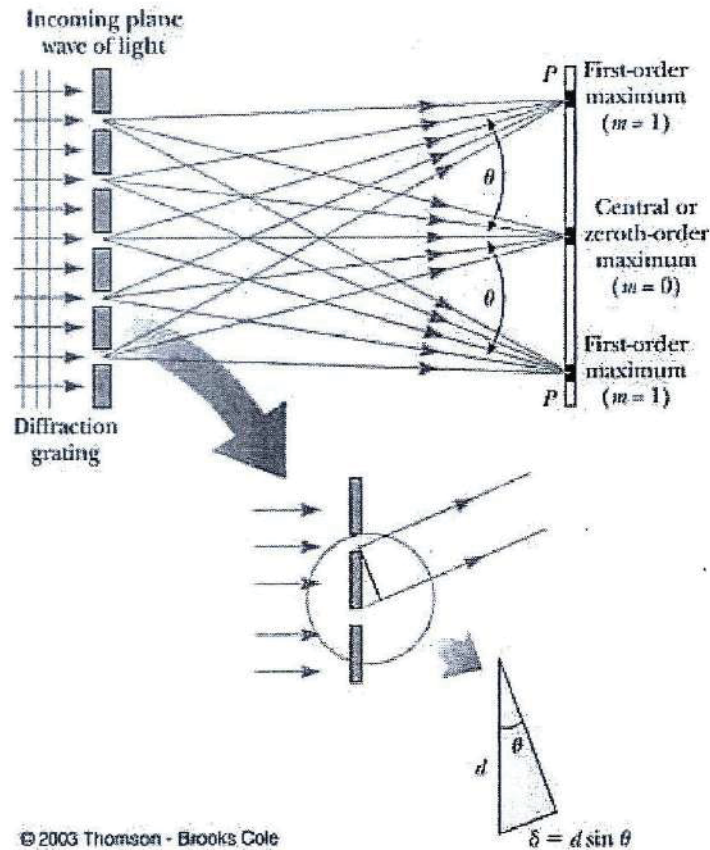


Fig. 10.7.1.

To see how a diffraction grating pattern differs from a Young's double-slit pattern and why it is so useful, we will analyse the diffraction pattern for various number of slits of the same slits-separation d .

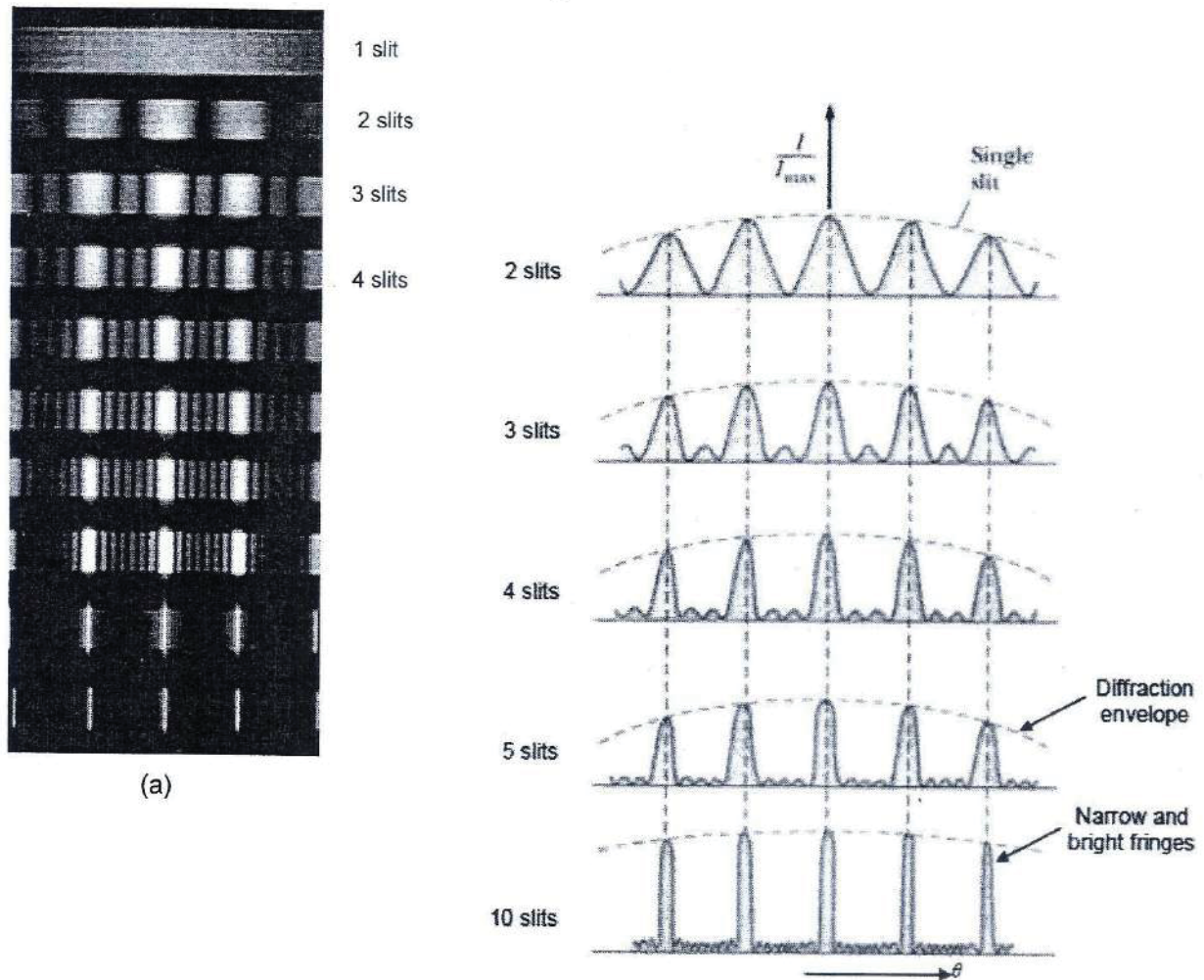


Fig 10.7.2. This shows the diffraction pattern of one slit, then two, then three, four, five, six, and more. When there are many slits, we call it a diffraction grating. The more slits, the narrower the pattern becomes. The pictures (a) above were taken with monochromatic light.

From the Fig. 10.7.2(a) above, we make the following observations:

- For the same slits separation d , the **positions of the maxima** ^(b) remain the same.
- As the number of slits increase, the maxima become narrower, i.e. the fringes become sharper on the screen.
- As the number of slits increase, the intensity of the maxima increases.

Therefore, we can infer that for a grating (no. of slits $\rightarrow \infty$) the **fringe pattern will have very good contrast** (narrower and brighter maxima separated by broad regions of dark minima).

Condition for Maxima for a Diffraction Grating

- From the above observations of the diffraction patterns of various slits, we can therefore infer the equation that determines the positions of the maxima for a *diffraction grating pattern*.
- For same value of d , the positions of the maxima (bright fringes) of a grating coincide with the positions of the maxima for a double slit :

Condition for Constructive Interference (Maxima) for a Diffraction Grating

For the n th order *maxima*,

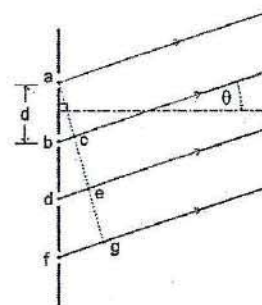
$$d \sin \theta = n\lambda$$

where d : separation between the slits

θ : angular position of the bright fringe from the principle axis

n : order number ($n=0, 1, 2, 3, \dots$)

λ : wavelength of light passing through the grating



Commercially available gratings are labelled by the *no. of lines per unit length*, N . The relationship between N and d are given by

$$N = \frac{1}{d}$$

Note :

- In this case, we *do not define* the conditions for *destructive interference* as the width of the region between the bright fringes is very large.
- For most of the gratings in the laboratory, the slit separations are much smaller than that of the double slits. Hence, you will notice that the positions of the first and higher orders of the bright fringes lie much further away from the central bright fringe and hence small angle approximation does not hold. Therefore, in most questions for diffraction grating it is not meaningful to talk about the *fringe separation* as the angular separation between consecutive fringes are not constant, in fact they *increase* as the order increases.



Although the formula for constructive interference for the double slit and the diffraction grating is the same, the formula for fringe separation for double slits, $\Delta y = \frac{\lambda L}{d}$, CANNOT be used in a grating, unless the fringes are shown to be regularly spaced.

3 Grating vs. Young's Double Slit

The grating is often the preferred choice over the double-slit for the following reasons. For a diffraction grating,

- Fringes are sharper and positions are more precisely determined. There is therefore less uncertainty in measuring the position of the fringes, and less uncertainty in the computation of wavelength.
- Typically the angular displacements are large, so percentage uncertainty in measuring the angular displacement becomes relatively lower. This translates to a relatively lower percentage uncertainty in measuring wavelength as well.

4 Determination of Wavelength Using Grating³

- Advantages and disadvantages of using the higher orders to determine the wavelength as compared to first order:

Using $d \sin \theta = m\lambda$, if we know the value of the slits separation d , we can generally determine the wavelength of the light used by measuring the value of θ for a particular order of fringe. There are both advantages and disadvantages in using the higher orders to determine this wavelength.

- Advantages :**

- angular displacement θ is larger, therefore there is less percentage uncertainty in measuring the angle.

- Disadvantages :**

- May not be as bright as the first order bright fringe.

Example 10.7.1 : Diffraction Grating

White light from a source passes through a filter which transmits only wavelengths of 400 nm to 600 nm. When the filtered light falls normally on a diffraction grating, light of wavelength 600 nm in one order of the spectrum is diffracted at the same angle, 30° , as the 400 nm light in the adjacent order. Find the number of lines per mm for the grating. [Ans 417 lines per mm]

Let 600nm light be the n th order. Then $(n+1)$ th order is the 400nm light.

$$d \sin \theta = n\lambda_1$$

$$d \sin \theta = (n+1)\lambda_2$$

Since d & θ is the same,

$$n\lambda_1 = (n+1)\lambda_2$$

$$n = 2$$

Then substitute $n=2$ into $d \sin \theta = n\lambda_1$

$$d = \frac{n\lambda_1}{\sin \theta} = \frac{2(600 \times 10^{-9})}{\sin 30^\circ} = 2.4 \times 10^{-6} \text{ m}$$

$$N = \frac{1}{d} = 4.17 \times 10^5 \text{ lines per m} = 417 \text{ lines per mm}$$

³ The spectrometer is a piece of equipment that is commonly used with the grating to determine the wavelength of light. Do read more about it in Appendix D (Pg. 48).

Example 10.7.2 : Maximum no. of Fringes Observable.

A monochromatic source of 495 nm is incident normally on a diffraction grating which has 500 lines per mm. How many diffraction lines can be observed on the screen? [Ans:9]

Strategy: The maximum angular deviation from the principle axis is 90° . Using this condition, we can determine the maximum order and hence find the theoretical maximum number of fringes observable on the screen.

Max theoretical angular displacement possible is 90°

$$d \sin \theta = n\lambda$$

$$d \sin 90^\circ = n_{\max} \lambda$$

Max. observable order,

$$n_{\max} = \frac{d}{\lambda} = \frac{1}{(500 \times 10^3)(495 \times 10^{-9})} = 4.04 = 4$$

Hence, max. no. of diffraction lines observable on the screen

$$= 2(4) + 1 = 9$$

10.8 Stationary Waves

Consider two identical wave trains moving in opposite directions overlapping in the same region of space. Figure 10.8.1 shows wave P moving towards the right and wave Q moving towards the left. By applying the principle of superposition, we know that during the time the two waves overlap, the resultant displacement is given by the vector sum of the displacements of the primary waves at that location.

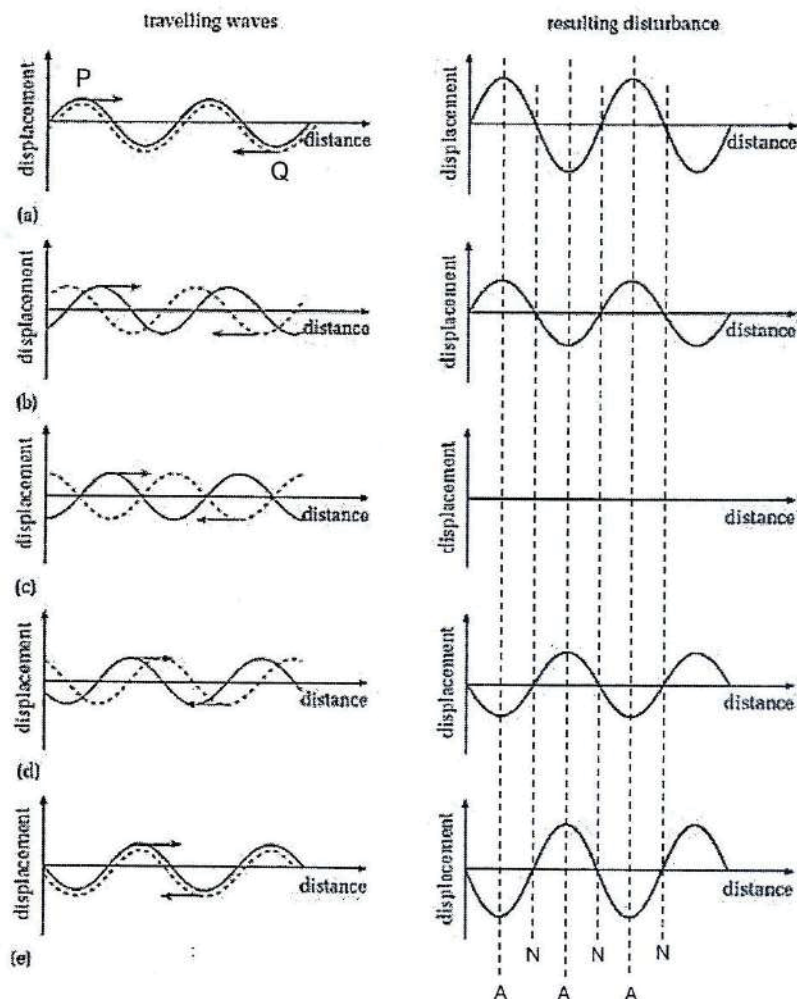


Fig. 10.8.1

The diagram shows how the superposition of 2 identical waves moving in opposite directions overlapping to form a stationary wave.

Because the wave pattern does not appear to move in either direction along the wave, the resultant waveform is known as a **standing wave** or **stationary wave**.

We can observe from the diagram that amplitude of vibration for every point of the resultant wave varies.

- There are particular points called **nodes** (N) that never move, i.e. the amplitude of vibration is zero.
- Midway between the nodes are points called **antinodes** (A), where the amplitude of vibration is greatest.

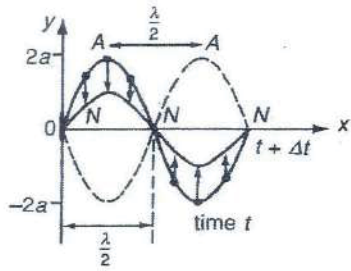
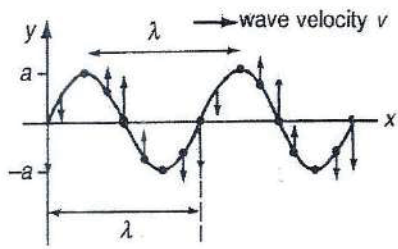
Formation of Stationary Wave (or Standing Wave)

Stationary Waves

When two identical waves of the **same amplitude, frequency and speed** but travelling in **opposite directions** are superposed together, the resultant wave obtained is called a stationary wave.

10.8.1 Characteristics of Stationary Wave

Let us examine the properties of the stationary wave by comparing it with a travelling wave.

	Standing Wave	Progressive Wave
Diagram		
Waveform	Does not advance.	Waveform advances with the velocity of the wave.
Amplitude	Varies according to position from zero at the nodes (permanently at rest) to a maximum of $2A$ at the antinodes. (A is the amplitude of each of the component travelling waves that form the stationary wave.)	Apart from attenuation, it is the same for all particles in the path of the wave.
Frequency	All the particles vibrate in SHM with the same frequency as the component waves (except for those at the nodes which are at rest).	All particles vibrate in SHM with the frequency of the wave.
Phase	Phase of all particles between 2 adjacent nodes is the same (in phase). Particles in adjacent segments between two nodes have a phase difference of π rad (in antiphase).	All particles within one wavelength have different phase.
Wavelength	2 x distance between a pair of adjacent nodes or antinodes.	Distance between two adjacent particles that are vibrating in phase.
Energy	No translation of energy, but there is energy associated with the wave.	Energy translates in the direction of the wave travel.

Example 10.8.1 : Stationary Waves

Fig 10.8.2 shows a car driven at a speed of 30 m s^{-1} along a straight road between 2 radio transmitters. The transmitters T_1 and T_2 are sending out the same programme, using a frequency of 1.50 MHz . The radio is heard to fade and strengthen regularly.

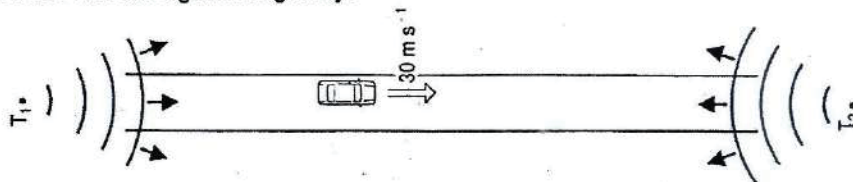


Fig. 10.8.2

What is the period of this regular fading? [Ans: 3.3 s]

As 2 sources are identical & transmit radio waves in opp. direction, a stationary wave will be formed b/w T_1 & T_2

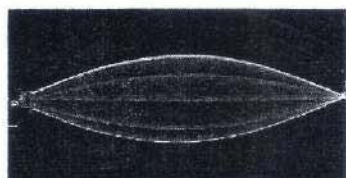
Period of the regular fading $\Rightarrow$ Time taken to move b/w 2 nodes
Dist. b/w 2 nodes $\Rightarrow \lambda/2$

$$\text{Thus, } \lambda \text{ of radio waves} = \frac{c}{f} = \frac{3.00 \times 10^8}{1.50 \times 10^6} = 200 \text{ m}$$

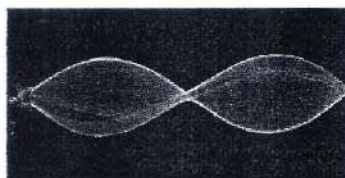
$$\text{Period of regular fading} = \frac{\text{Dist. b/w 2 nodes}}{\text{Speed of car}} = \frac{200/2}{30} = 3.3 \text{ s}$$

10.8.2 Stationary Waves in Strings

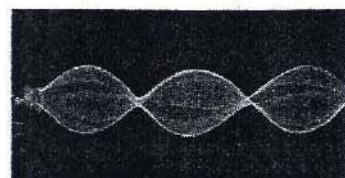
Consider a string that is fixed at one end. Suppose its free end is moved up and down in simple harmonic motion, the oscillation will produce a continuous wave-train down the string. As the wave-train reaches the fixed end, it gets reflected back. The reflected wave-train will interfere with the wave-train from the source to form a resultant waveform on the string. It is observed that at certain frequencies of vibration of the source, we will be able to see some unique waveforms on the string as illustrated in Fig. 10.9.3(a)-(c).



(a)



(b)



(c)

Fig. 10.8.3

Time exposures of standing waves in a stretched string. At certain driven frequencies of vibration, the amplitude of vibration of the string become very large and the resultant waveform does not appear to progress along the string. (a) to (c) show standing waveforms at increasing frequencies of vibration.

The frequencies at which standing waves are produced are known as the **natural frequencies** or **resonant frequencies** of the string.

The different standing wave patterns shown in Fig. 10.8.3 are the different "**resonant modes of vibration**". When the driving frequency matches that of the natural frequency of the string, the amplitude of oscillation generated are the largest.

This principle is very important as it is used in the generation of music in many stringed instruments like guitar and violin.

Resonant modes of vibrations in a stretched string fixed at two ends

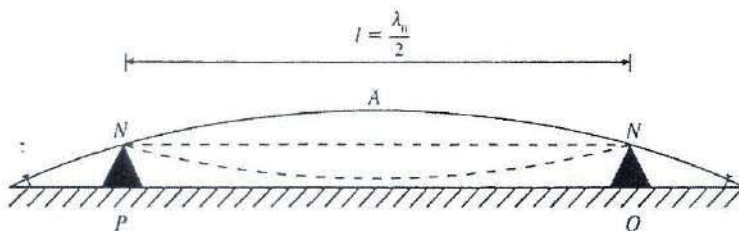


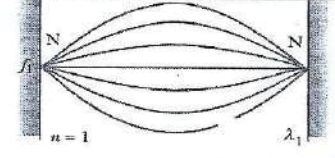
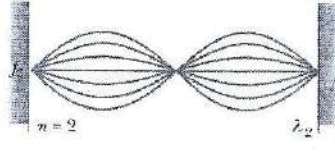
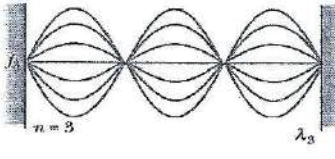
Fig. 10.8.4

Consider a wire stretched between 2 points and plucked in the middle, a transverse wave travels along the wire and is reflected at the ends. The reflected and re-reflected waves from the ends will continually interfere with each other to produce a stationary wave on the string.

The two ends of the strings are fixed and thus must have zero displacement and hence must be displacement nodes.

This additional constraint or boundary condition results in the string being allowed to have only certain wavelengths and frequencies of oscillations.

Fig. 10.8.4 shows the largest possible wavelength of stationary wave that can be produced on the wire - corresponding to the smallest frequency. This mode of oscillation is known as the **fundamental mode** and the lowest frequency of vibration f_1 is known as the **fundamental frequency**.

Mode of Vibration	Wavelength	Frequency	Harmonic	Overtone
 $n = 1$	$\lambda_1 = 2L$	$f_1 = \frac{v}{2L}$	1 st	—
 $n = 2$	$\lambda_2 = L$	$f_2 = \frac{v}{\lambda_2}$ $= \frac{v}{L} = 2\left(\frac{v}{2L}\right)$	2 nd	1 st
 $n = 3$	$3\left(\frac{\lambda_3}{2}\right) = L$ $\lambda_3 = \frac{2}{3}L$	$f_3 = \frac{v}{\lambda_3}$ $= \frac{3v}{2L} = 3\left(\frac{v}{2L}\right)$	3 rd	2 nd

Generally the n th harmonic is given by: $f_n = n\left(\frac{v}{2L}\right)$, where $n = 1, 2, 3, 4, \dots$

Note:

1. Fundamental Frequency and Harmonics

All resonant frequencies of vibrations that can be generated are called harmonics. The harmonics are numbered according to the ratio of their frequencies to the first natural frequency f_1 . Thus if the 1st harmonic is 110 Hz, then 220 Hz, 330 Hz, 440 Hz etc. are the 2nd, 3rd, 4th etc harmonics respectively.

- The first natural frequency is the lowest possible frequency and is known as the *fundamental frequency* (or 1st harmonic)
- The frequencies of the harmonics are related to the fundamental frequency f_1 by $f_n = nf_1$ where f_n is the frequency of the n th harmonic.
- Note that all harmonics are present for the string shown above.

2. Overtones

Overtones are the frequencies (excluding the fundamental) that can be produced by an instrument accompanying the 1st harmonic that is played.

The tuning fork and frequency generator are the only examples that can produce a pure note.

The lowest overtone is known as the 1st overtone, the next overtone is the 2nd overtone. Hence, in this case, the 1st overtone of the string corresponds to the 2nd harmonic of the string.

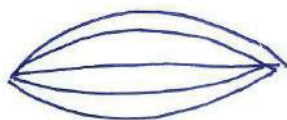
3. Speed of a wave on a string

The speed of a wave on a string is given by $v = \sqrt{\frac{T}{\mu}}$ where T is the tension in the string and μ is the mass per unit length of the string.

Example 10.8.2: Resonant Wavelengths of a Wire

A 1.0 m wire stretched between 2 points is plucked near one end. What are the three longest wavelengths present on the vibrating wire?

[2.0 m, 1.0 m, 0.67 m]



$$\lambda_1 = 2L = 2.0 \text{ m}$$



$$\lambda_2 = L = 1.0 \text{ m}$$



$$\lambda_3 = \frac{2}{3}L = 0.67 \text{ m}$$

10.8.3 Standing Waves in Air Column

Like standing transverse wave on a stretched string, a standing longitudinal wave is obtained by the *superposition of two trains of progressive longitudinal waves of equal amplitude, and frequency, travelling with the same speed in the opposite directions.*

Stationary transverse and longitudinal waves share similar properties except for the *direction of the displacement of the particle* where the displacement of air molecules are parallel to the direction of propagation of the component waves that form the stationary wave.

The formation of stationary waves in pipes and tubes is the basis of production of musical notes in most wind instruments. When stationary waves are set up in the tubes, **the air column resonates at its natural frequencies.**

The boundary conditions that determine the natural frequencies of the air column depends on whether that end is open or closed.



Fig. 10.8.5

Looking at Fig. 10.8.5,

At the closed end,

- the longitudinal movement of the air molecules is prevented and is always a **displacement node**.

At the open end,

- air molecules can vibrate freely, and is always a **displacement antinode**.

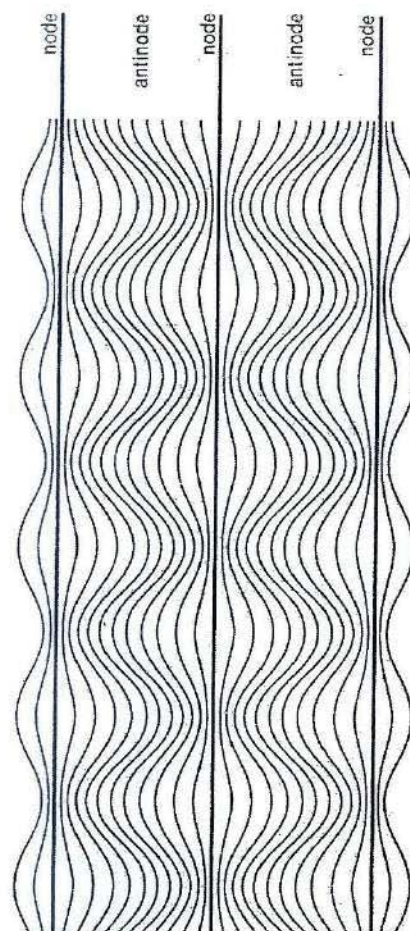


Fig. 10.8.6

The diagram can be used to show a longitudinal standing wave. To see the effect, move a piece of stiff card with a slot of no more than 1 mm cut in down over the diagram.

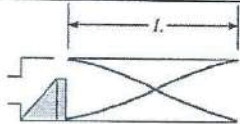
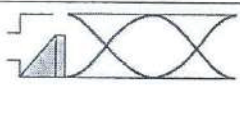
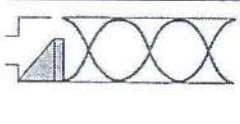
Resonant Frequencies of Air Columns

There are generally 2 types of air columns we want to concern ourselves with :

- open pipe
- closed pipe (closed at one end)

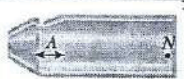
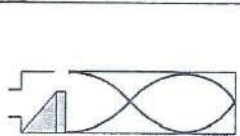
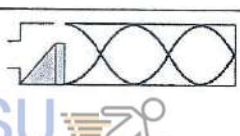
Recall that the open end of the pipe, we have a displacement antinode, and at the closed end of the pipe, we have a displacement node.

Resonant Vibrations in an Open Pipe

	(Displacement) Wave Profile	Wavelength	Frequency	Harmonic	Overtone
		λ_1	f_1	1 st (fundamental)	—
		$\lambda_2 = L$	$f_2 = \frac{v}{\lambda_2}$ $= \frac{v}{L} = 2\left(\frac{v}{2L}\right)$	2 nd	1 st
		$\frac{3}{2}\lambda_3 = L$ $\lambda_3 = \frac{2}{3}L$	$f_3 = \frac{v}{\lambda_3}$ $= \frac{3v}{2L} = 3\left(\frac{v}{2L}\right)$	3 rd	2 nd

The n th harmonic can be expressed as: $f = n\left(\frac{v}{2L}\right)$, where $n = 1, 2, 3, 4, \dots$

Resonant Vibrations in a Closed Pipe

	(Displacement) Wave Profile	Wavelength	Frequency	Harmonic	Overtone
		λ_1	f_1	1 st (fundamental)	—
		$\frac{3}{4}\lambda_3 = L$ $\lambda_3 = \frac{4}{3}L$	$f_3 = \frac{v}{\lambda_3}$ $= \frac{3v}{4L} = 3\left(\frac{v}{4L}\right)$	3 rd	1 st
		$\frac{5}{4}\lambda_5 = L$ $\lambda_5 = \frac{4}{5}L$	$f_5 = \frac{v}{\lambda_5}$ $= \frac{5v}{4L} = 5\left(\frac{v}{4L}\right)$	5 th	2 nd

The n th harmonic is given by: $f_n = n\left(\frac{v}{4L}\right)$, where $n = 1, 3, 5, \dots$

Note:

1. The fundamental frequency of an open pipe is twice that of a closed pipe of the same length.
2. Only odd harmonics are present in closed pipe.
3. Assuming that we have a closed pipe and an open pipe with characteristics the same in all aspects inclusive of length, the following table presents a possible situation for the combination of frequencies emitted.

Harmonics	Frequency emitted from a Closed Pipe /Hz	Frequency emitted from an Open Pipe / Hz
1	256	512
2	-	1024
3	768	1536
4	-	2048
5	1280	2560

The quality of sound emitted is a mixture of these harmonics for each system. Hence we see that the "note" from an open pipe is "richer" than that from a closed pipe.

Example 10.8.3 Serway p18.37.

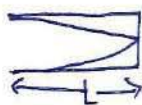
Calculate the length of a pipe that has a fundamental frequency of 240 Hz if the pipe is

- (a) closed at one end and
- (b) open at both ends.

Take the speed of sound to be 343 m s^{-1} .

[Ans: 0.357m, 0.715 m]

a) For pipe w 1 end closed and vibrating at fundamental freq,



$$\begin{aligned}
 L &= \lambda/4 \\
 &= \frac{v}{4f} \\
 &= \frac{343}{4 \times 240} = 0.357 \text{ m}
 \end{aligned}$$

b) For pipe open at both ends & vibrating at fundamental freq,



$$\begin{aligned}
 L &= \lambda/2 \\
 &= \frac{v}{2f} \\
 &= \frac{343}{2 \times 240} = 0.715 \text{ m}
 \end{aligned}$$

Example 10.8.4 Serway P18.45.

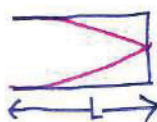
An air column in a glass tube is open at one end and closed at the other by a movable piston. The air in the tube is warmed above room temperature, and a 384-Hz tuning fork is held at the open end. Resonance is first heard when the piston is 22.8 cm from the open end and again when it is 68.3 cm from the open end.

- (a) What speed of sound is implied by these data?
 (b) How far from the open end will the piston be when the next resonance is heard?
 [(a) 350 m s⁻¹ (b) 1.14 m]



a) when piston is 22.8 cm from the open end, length L of pipe is $\frac{1}{4}\lambda$

$$\begin{aligned} v_1 &= f\lambda \\ &= f(4L) \\ &= (384)(4 \times 22.8 \times 10^{-2}) \\ &= 350.2 \text{ m/s} \end{aligned}$$



when piston is 68.3 cm from the open end, length L_3 of pipe is $\frac{3}{4}\lambda$

$$\begin{aligned} v_2 &= f\lambda \\ &= f(4L_3/3) \\ &= (384)(4 \times 68.3 \times 10^{-2})/3 = 349.7 \text{ m/s} \end{aligned}$$



Speed of sound,
 $v = (350.2 + 349.7)/2$

End Corrections⁴

In practice, the air just outside the open end of a pipe is set into vibration and the displacement antinode of a standing wave occurs a distance c – called the **end correction** beyond the end of the tube.

The effective length of the resonant air column is larger than the length of the pipe. The following figure shows the respective corrected lengths for the fundamental frequencies of a closed and open pipe.

Closed Pipe	Open Pipe
Making Amendments for End Correction $\frac{\lambda}{4} = L + c$	Making Amendments for End Correction $\frac{\lambda}{2} = L + 2c$

- Note that the end correction for a cylindrical pipe is approximately $0.6r$ where r is the radius of the pipe. Hence, for the larger pipes, it is necessary to take end correction into consideration.

⁴ The knowledge of end correction is not stipulated in the syllabus. But this can be found in many questions in A-levels especially in the earlier years.

10.8.4 What about the pressure?

Earlier we discussed the pressure variations in a progressive sound wave; now we can consider how pressure varies in a stationary sound wave.

Fig. 10.8.7 (a) shows the two extreme curves of the displacement-position graph for a stationary wave for times when the displacements are greatest. As before, we will take displacements of air molecules to the right to be positive.

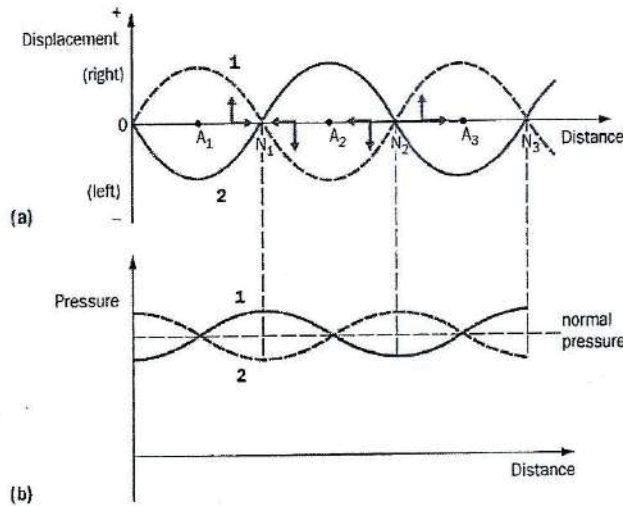


Fig. 10.8.7

Pressure variation for a stationary sound wave

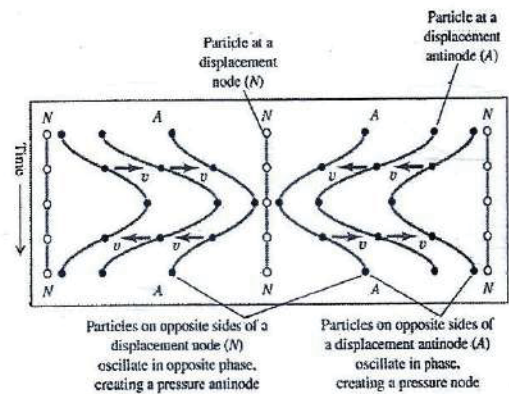


Fig. 10.8.8

In a standing longitudinal wave, a displacement node N is a pressure antinode (a point where the pressure fluctuates the most) and a displacement antinode A is a pressure node (a point where pressure does not fluctuate).

Consider graph (a),

- Air to the left of node N_1 is displaced positively and hence to the right towards N_1 ; air to the right of N_1 has a negative displacement and hence to the left toward N_1 as well.

At N_1 , the air is compressed and the pressure is greater than normal.

- Hence, we can conclude

Pressure variation is greatest at displacement nodes.

- Therefore, for a longitudinal stationary wave, we can conclude that :

Displacement node is subjected to large pressure changes and hence it is a **pressure antinode**.

- Since nodes are always located at midpoints of 2 adjacent antinodes, therefore

Displacement antinodes corresponds to points of **pressure nodes**

At those points, molecules are free to move about and hence the pressure remains at normal pressure.

- Now consider N_2 , in this case the air molecules to the left and right of N_2 all moves away from N_2 . Therefore, at N_2 , the air is less dense than usual.

At N_2 , the air is rarefied and the pressure is lower than normal.

KIASU
ExamPaper

Example 10.8.5: End Corrections

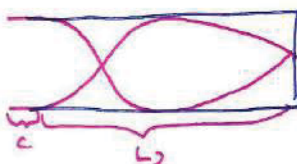
A source of sound of frequency 250 Hz is used with resonant tube, closed at one end, to measure the speed of sound in air. The resonant lengths are found to be 0.30 and 0.96 m. Calculate

- (a) the speed of the sound, and
 (b) the end correction of the tube.

[Ans: (a) 330 m s⁻¹ (b) 3.0 cm]

a) 

$$\frac{\lambda}{4} = L_1 + c \quad \text{--- ①}$$



$$\frac{3\lambda}{4} = L_2 + c \quad \text{--- ②}$$

$$\text{②} - \text{①} : \frac{\lambda}{2} = L_2 - L_1$$

$$\therefore \text{speed of sound, } v = 2(250) \frac{(0.96 - 0.30)}{1} \\ = 330 \text{ m/s}$$

b) $\frac{\lambda}{4} = L_1 + c$

$$c = \frac{\lambda}{4} - L_1 = \frac{v}{4f} - L_1 = \frac{330}{4(250)} - 0.30 = 0.03 \text{ m} \\ = 3 \text{ cm}$$

10.8.5 Investigating the speed of sound using stationary waves

We can determine the speed of sound wave by setting up a stationary wave and measure the corresponding frequency and wavelength as described by the experiment below. A stationary sound wave is set up by reflection from a reflector.

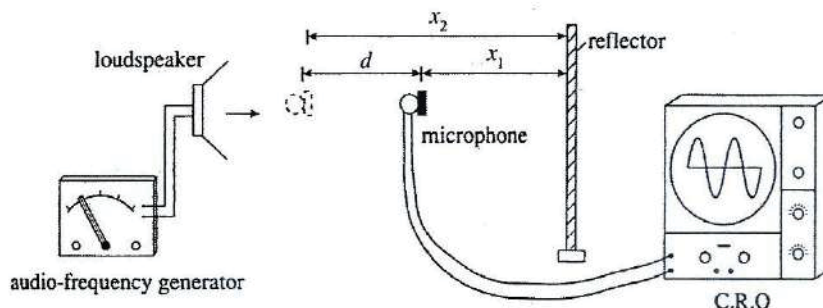


Fig. 10.8.9 Measurement of speed of sound

The experimental setup is shown in Fig. 10.8.9.

- The signal generator (See Appendix E, Pg. 49) connected to a speaker delivers a note of frequency f towards a reflector (e.g. a metal sheet).
- The sound wave gets reflected off the reflector and interference occurs between the forward and reflected waves and a stationary wave pattern with nodes and antinodes is established between the speaker and reflector.
- The positions of antinodes are determined using a microphone connected to the oscilloscope (See Appendix E, Pg. 49). The microphone is slowly moved away from the reflector until a trace of maximum amplitude is obtained. This denotes the position of an antinode. The distance x_1 of the microphone from the reflector is measured using a metre rule.
- The microphone is next moved slowly further away from the reflector until the amplitude of the trace on the C.R.O. is once again maximum. The distance of the microphone from the reflector, x_2 , is again measured.

Therefore, Distance between 2 successive antinodes, $d = x_2 - x_1 = \lambda/2$

Wavelength of the sound wave, $\lambda = 2d$

- If the frequency f of the sound wave is known, the speed of sound v can be determined from

$$v = f\lambda$$

Appendices

Appendix A: Huygens' Principle

Huygens' Principle

But how can one explain the bending of waves due to diffraction? Why does the spreading of waves become increasingly more obvious with the width of the aperture decreases to about the same order as the wavelength? This phenomenon may be explained using *Huygens' Principle*.

The *Huygens' Principle* is a geometrical method for finding, from the known shape of a wave front at some instant, the shape of the wave front at some later time. Huygen assumed that

every point of a wave front may be considered the source of secondary wavelets that spread out in all directions with a speed equal to the speed of propagation of the wave.

The new wave front at a later time is found by constructing a surface *tangent to the secondary wavelets*, called the *envelope* of the wavelets.

The following diagram shows how Huygens' construction is used to determine the new wavefronts for a plane wave and for a spherical wave.

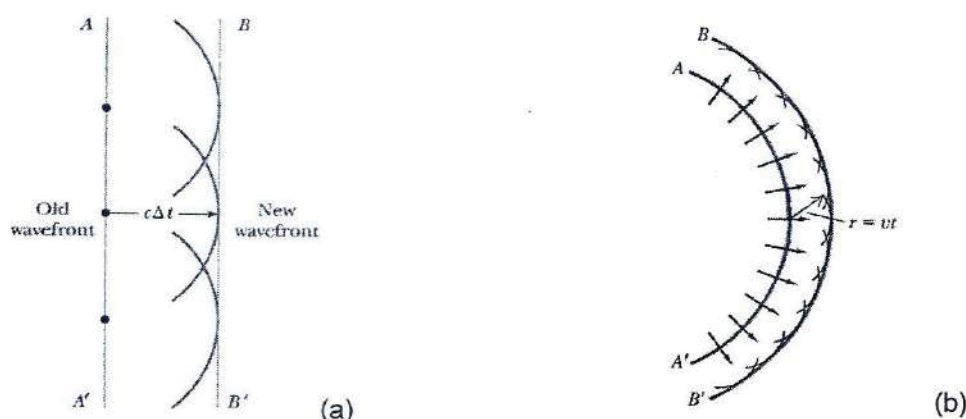
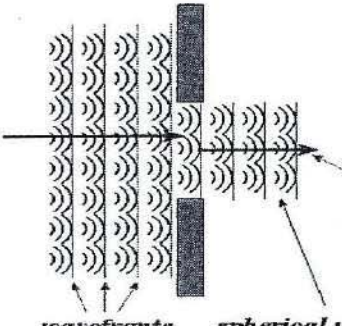
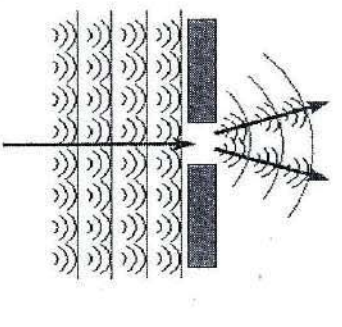
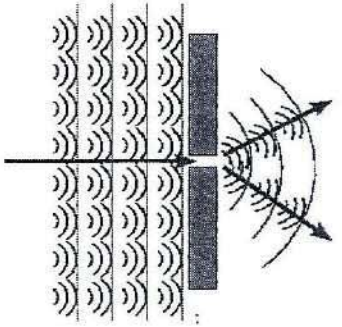


Fig. 1: Huygens' construction for (a) a plane wave propagating to the right and (b) a spherical wave propagating to the left.

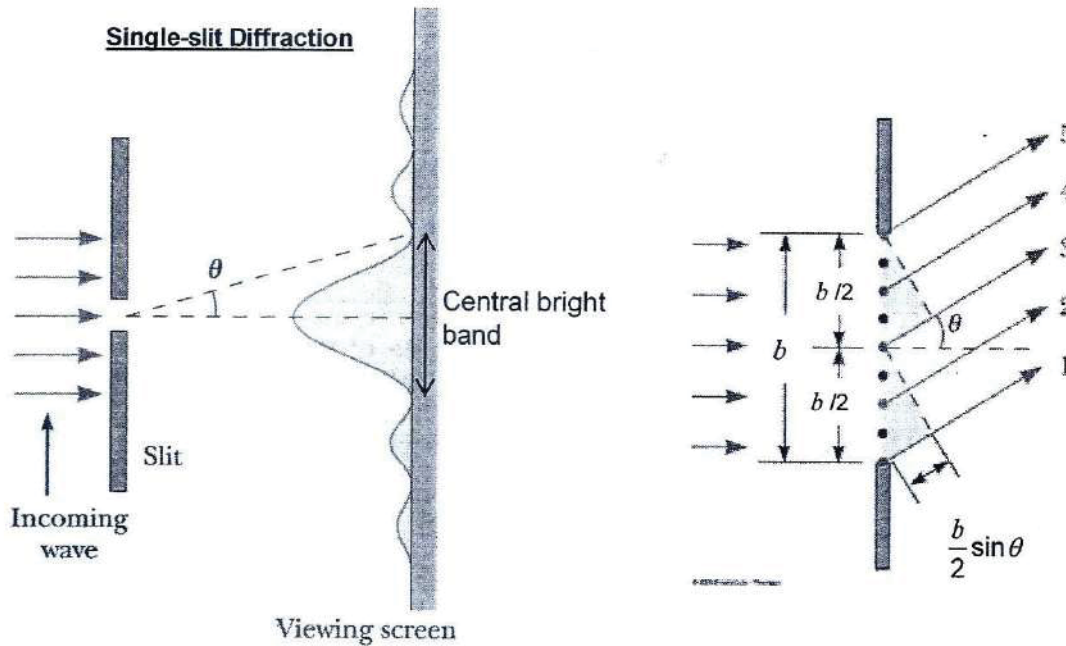
Through Huygens' construction, we can now see how the subsequent wavefronts 'reconstruct' themselves when incident plane waves pass through a slit.

Let us now apply Huygens' construction to explaining diffraction.

 Fig. 2	Consider first the case where the obstacle or opening is much larger than the wavelength of the waves (Fig. 2), we find that the wavelets emerging forms wavefronts that is essentially straight. Diffraction is not very apparent.
 Fig. 3	If, however, the size of the opening becomes comparable to the wavelength, the waves proceed to "bend through" or around the opening or obstruction as shown in Fig. 3.
 Fig. 4	Finally, as the obstruction or opening equals the wavelength, the diffraction becomes quite pronounced as shown in Fig 4.

Appendix B: Single Slit Diffraction Pattern

To analyze the diffraction pattern of the single slit, let us divide the slit into two halves as shown.



Consider ray 1 and 3. As the two rays travel towards the screen, ray 1 travels an additional path difference of $\frac{b}{2} \sin \theta$ where b is the width of the slit and θ is the angle with the principle axis. Similarly, the path difference between rays 2 and 4, as well as between rays 3 and 5, is also $\frac{b}{2} \sin \theta$.

If this path difference is exactly half a wavelength, then destructive interference occurs. This cancellation occurs for any two rays that originate at the points separated by half a slit width because the phase difference between 2 such points is 180° . Therefore the waves from the upper half of the slit interfere destructively with the waves from the lower half when

$$\frac{b}{2} \sin \theta = \pm \frac{\lambda}{2}$$

$$\text{or } b \sin \theta = \pm \lambda$$

Dividing the slit into 4 equal parts and using the same reasoning, we find that destructive interference also occurs when $\frac{b}{4} \sin \theta = \pm \frac{\lambda}{2}$ or $b \sin \theta = \pm 2\lambda$.

This same argument can be applied when the slit is divided into 6, 8 and so on equal parts, and we obtained the general condition for **destructive interference** (or **dark region**) to be

$$b \sin \theta = m\lambda \text{ where } m = \pm 1, 2, 3, \dots$$

Appendix C: Ripple Tank

One way to observe interference of water waves is in a ripple tank. Generally a ripple tank could be used to investigate various properties of water waves.

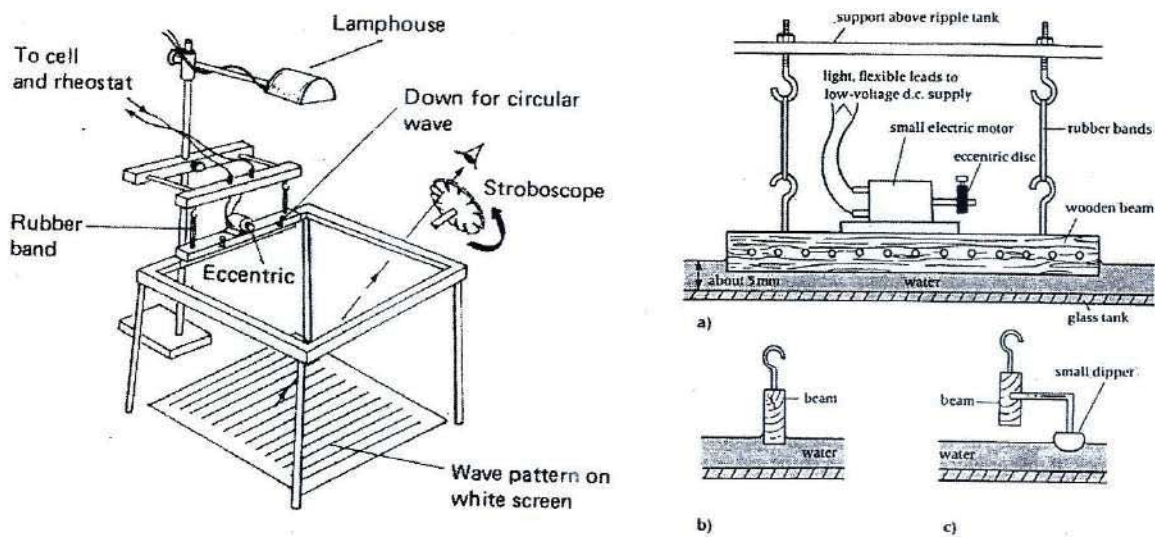


Fig. 1 Set-up for the ripple tank

1. **Eccentric:** An eccentric (off-centre) metal disc on the axle of the motor causes the beam to vibrate at the same frequency as the revolutions of the motor. The speed of the motor is controlled by a rheostat in series with a low-voltage d.c. supply, or a dry battery (usually 3.0 or 4.5 V).
2. To generate continuous straight wave, a vibrating beam just touching the water is used.
3. To generate circular waves, a small spherical dipper is fixed to the beam instead.
4. The speed of the motor may be adjusted to vary the frequency and wavelength of continuous wave train.
5. The progressive waves move quite quickly across the water surface and any attempt to measure the wavelength or draw the wave diagrams from the moving water shadows would be very difficult. A standard method of freezing or slowing down rapid motion would be to use a stroboscope.

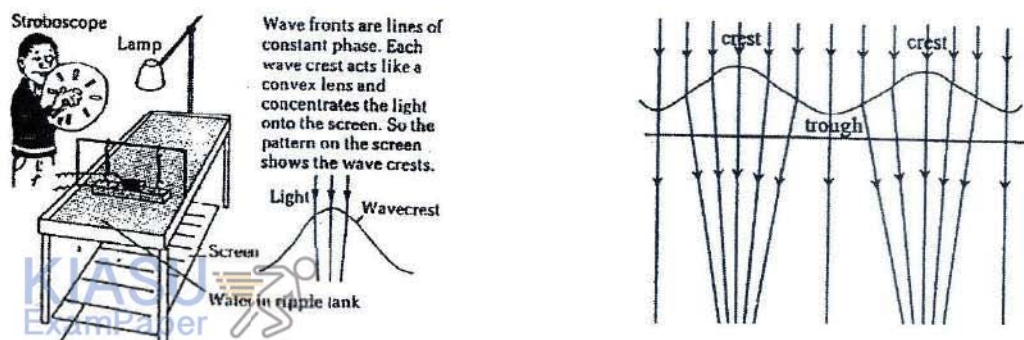


Fig. 2 Observations of wavefronts in a ripple tank.

Appendix D: Spectrometer

The grating or double slit is usually mounted on a device known as a *spectrometer* to allow us to easily take measurements of θ so that we can study about the light source. Fig. 1 below shows you a typical laboratory spectrometer and Fig. 2 shows a schematic sketch of the spectrometer.

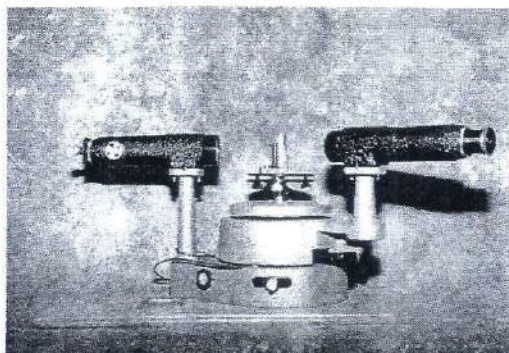


Fig. 1
A laboratory spectrometer

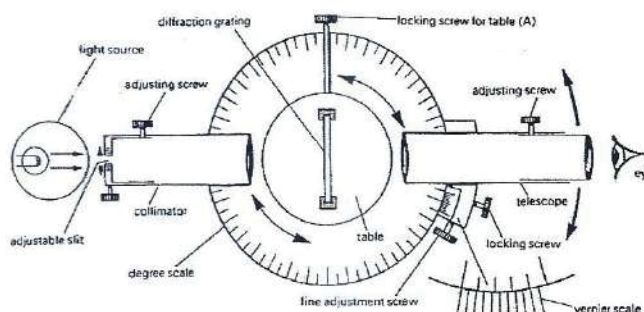


Fig. 2
The schematics of the spectrometer

The essential parts of a spectrometer are

- A fixed *collimator* with a movable slit of adjustable width to produce a parallel beam of light from the source illuminating the slit. The light source to be studied is placed between the collimator and the turntable.
- a *turntable* on which the prism, double slit or grating can be mounted, and
- a *telescope* attached to the turntable such that it can be rotated about same vertical axis as the turntable. Usually there is also vernier scale attached below the turntable so that you to read off the *angle* through which the telescope is turned through from this scale. The vernier scale allows you to read the angular position of the telescope to the nearest 0.1° .

When used with a grating,

- we usually start off by locating the position of the *zeroth order maximum*. It is the brightest fringe that has the brightest intensity. This provides a reference point for which angular displacement can be measured. The angular displacement of the m^{th} order maximum,

$$\theta = |\text{Angular position of zeroth order bright fringe} - \text{Angular position of } m^{\text{th}} \text{ order bright fringe}|$$

$$\Rightarrow \theta = |\theta_o - \theta_m|$$

Alternatively, by noting the angular positions of the m^{th} order bright fringe at either side of the central maximum ($\theta_{m,\text{left}}$ and $\theta_{m,\text{right}}$), we can also determine the angular displacement θ of m^{th}

order maximum from the central maximum. In this case, $2\theta = |\theta_{m,\text{left}} - \theta_{m,\text{right}}|$

- We also move the telescope to each side of this central maximum to check that several orders of *diffraction* can be seen and there is some symmetry about this central maximum. If not, the plane of the surface of the grating is not perpendicular to the plane in which telescope moves and tilt of the table needs to be adjusted.

Appendix E: Function Generator and C.R.O.

Function Generator

A function generator is an instrument that produces alternating (a.c.) voltage of different frequencies.

The standard voltage signals it can produce are:

- sine waves,
- triangular waves
- rectangular waves

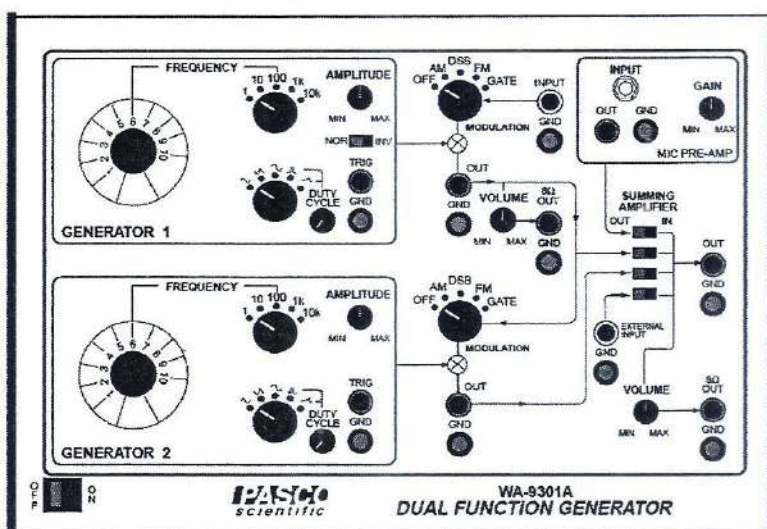
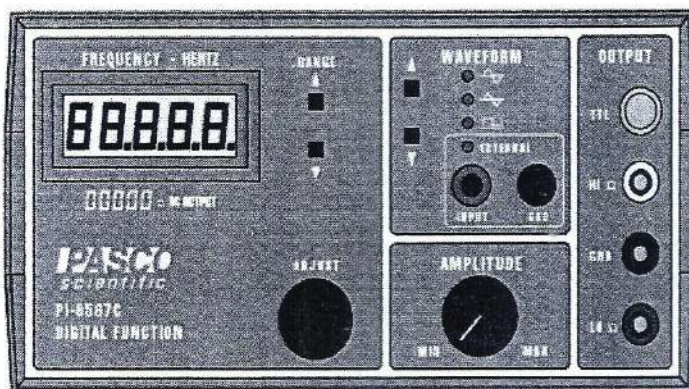


Fig. 1. Examples of function generators. The top one is a single output generator. The bottom generator is able to produce two outputs at the same time.

As not all function generators have a precise digital display of the frequency of the waveform or the amplitude of the voltage wave produced. It will sometimes be necessary to connect the function generator to a CRO in order to determine the precise frequency and amplitude of the voltage wave.

To generate the sound waves, the function generator is often connected to a speaker so that the speaker can convert the voltage signals to vibrations of the membrane of the speakers to produce sound waves.

Cathode Ray Oscilloscope

The other commonly used equipment in sound experiments is the **cathode ray oscilloscope (CRO)**. The oscilloscope is basically a graph displaying device that traces the variation of the input voltage signal with respect to time.

Using the oscilloscope, you can measure the amplitude, period, and frequency of any voltage signal. Therefore, a CRO is actually a more complex voltmeter. Hence, to measure the frequency of a sound wave, we need to first connect a microphone to the scope. The microphone will then convert the vibrational energy of the air molecules to voltage signals which can be measured by the CRO.

Most scopes can display two signals on the screen at one time, enabling you to observe their time relationships. These scopes are known as dual-channel oscilloscopes. A typical front panel view of such a scope is shown in Fig. 10.12.2.

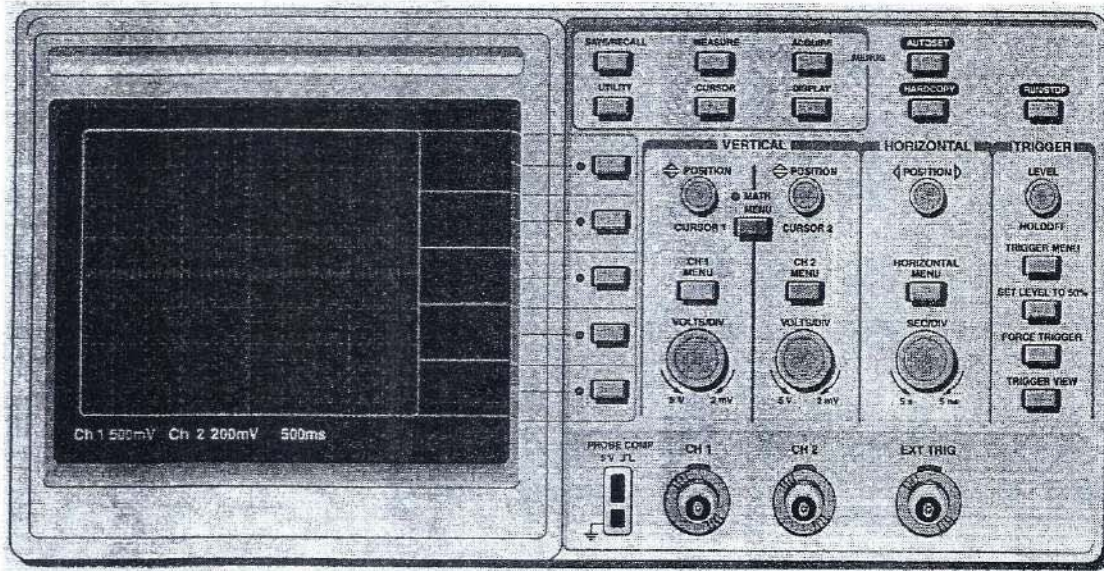


Fig. 2. A typical dual channel oscilloscope. Numbers below screen indicate the values for each division on the vertical (voltage) and horizontal (time) scales and can be varied using the vertical and horizontal controls on the scope.

Actual CROs vary depending on model and manufacturer, but most have certain common features. The main controls will be discussed below. You will need to refer to the user manual for complete details of the scope you are using.

Vertical Controls:

In the vertical section of the scope (Fig. 2), there are identical controls for each of the two channels (CH1 and CH2).

- **Position Control.** The position control lets you move a displayed waveform up or down vertically on the screen.
- **Volts/Div control.** The Volts/Div control is also known as the **Y-sensitivity / Y-gain control**. It adjusts the number of volts represented by each vertical division (usually 1 cm) on the screen. For this scope, the Volts/Div setting for each channel is displayed at the bottom of the screen. For CH1, it is 500 mV/div and for CH2, it is 200 mV/div.

Horizontal Controls:

In the horizontal section, the controls apply to both channels.

- **Position Control.** The position control lets you move a displayed waveforms left or right horizontally on the screen.
- **Sec/Div Control.** The Sec/Div control is also known as the **Timebase** control, it adjusts the time represented by each horizontal division or main time base. Similarly, for this scope, the Sec/Div setting is displayed at the bottom of the screen.

TOPIC 11 THERMAL PHYSICS

H2 Physics 9749 Syllabus (Examined from 2017)

Temperature and Ideal Gases

Content

- Thermal equilibrium
- Temperature scales
- Equation of state
- Kinetic theory of gases
- Kinetic energy of a molecule

Learning Outcomes

Candidates should be able to:

- (a) show an understanding that regions of equal temperature are in thermal equilibrium
- (b) explain how empirical evidence leads to the gas laws and to the idea of an absolute scale of temperature (i.e. the thermodynamic scale that is independent of the property of any particular substance and has an absolute zero)
- (c) convert temperatures measured in kelvin to degrees Celsius: $T / K = T / ^\circ\text{C} + 273.15$
- (d) recall and use the equation of state for an ideal gas expressed as $pV = nRT$, where n is the amount of gas in moles
- (e) state that one mole of any substance contains 6.02×10^{23} particles and use the Avogadro number $N_A = 6.02 \times 10^{23} \text{ mol}^{-1}$
- (f) state the basic assumptions of the kinetic theory of gases
- (g) explain how molecular movement causes the pressure exerted by a gas and hence derive the relationship $pV = (1/3) Nm\langle c^2 \rangle$, where N is the number of gas molecules (a simple model considering one-dimensional collisions and then extending to three dimensions using $1/3 \langle c^2 \rangle = \langle c_x^2 \rangle$ is sufficient)
- (h) recall and apply the relationship that the mean kinetic energy of a molecule of an ideal gas is proportional to the thermodynamic temperature (i.e. $\frac{1}{2} m\langle c^2 \rangle = (3/2) kT$) to new situations or to solve related problems.

First Law of Thermodynamics

Content

- Specific heat capacity and specific latent heat
- Internal energy
- First law of thermodynamics

Learning Outcomes

Candidates should be able to:

- (a) define and use the concepts of specific heat capacity and specific latent heat
- (b) show an understanding that internal energy is determined by the state of the system and that it can be expressed as the sum of a random distribution of kinetic and potential energies associated with the molecules of a system
- (c) relate a rise in temperature of a body to an increase in its internal energy
- (d) recall and use the first law of thermodynamics expressed in terms of the increase in internal energy, the heat supplied to the system and the work done on the system.

11.1 Internal Energy, Temperature and Heat

You are familiar with energies due to the bulk movement or position of a system as a whole. Internal energy, on the other hand, refers to microscopic energies due to the movement and interaction of individual constituents within the system.



For example, in the macroscopic view, a tennis ball flying across the court possesses kinetic energy of motion and potential energy of gravitation.

In the microscopic view, the molecules that make up the tennis ball also possess kinetic energy since they are in constant motion vibrating within the lattice. They also possess potential energy since they are held together by the intermolecular forces which are electrical in nature. These make up the internal energy of the tennis ball.

The **internal energy** is the summation of microscopic kinetic energy due to random motion of the gas molecules and the microscopic potential energy due to intermolecular forces.

$$U = \sum PE_{\text{microscopic}} + \sum KE_{\text{microscopic}}$$

Notes:

1. The internal energy of a system is an energy term and hence is measured in joules (J).
2. The commonly used symbol for representing internal energy is U .

It turns out that the temperature of a body is a measure of the **average kinetic energy** of the molecules that make up the body. Increases in internal kinetic energy of a body can be detected by a rise in its temperature.

Increases in internal potential energy, on the other hand, does not result in any temperature rise. This is most dramatically demonstrated by boiling water: despite all the heat being supplied, the thermometer continues to read 100 °C.

Thermal equilibrium

Two objects are said to be in thermal equilibrium when they are in thermal contact and there is **no net exchange of thermal energy** between them.

Temperature, T

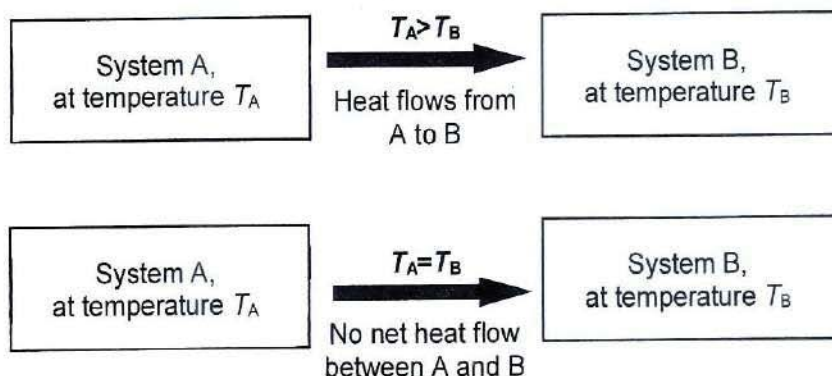
If two objects are in **thermal equilibrium**, they have the **same temperature**.

Heat, Q

Heat is defined as the **transfer of energy** between a system and its environment due to a difference in temperature between them.

Notes:

1. Heat is an energy term and hence is measured in joules (J).
2. The commonly used symbol for representing heat is Q .

**Zeroth Law of Thermodynamics**

If bodies A and B are **separately in thermal equilibrium** with a third body C, then A and B are also in thermal equilibrium with each other.

If the Zeroth Law looks like a no-brainer to you, you are not alone. Let's take a little time to examine this framework that we have cobbled together. A few interesting deductions:

- (1) It is temperature, not internal energy, which decides the direction of heat flow between two bodies. Thermal equilibrium is attained when the molecules in both bodies have the same average kinetic energy.
- (2) Two bodies at the same temperature may have different internal energy. For example, for the same temperature, 1 kg of water has less internal energy than 2 kg of water, simply because it contains fewer molecules. For another example, 1 kg of water at 0°C has more internal energy than 1 kg of ice also at 0°C , simply because it possesses higher potential energy.
- (3) A rise in temperature of a body indicates a rise in its internal energy. However, a rise in its internal energy may not lead to a rise in its temperature. (This will be elaborated on in the sections on Latent Heat)
- (4) Heat is the transfer of thermal energy. So it is wrong to say "a water bath can store a lot of heat", or "a hot object has a lot of heat".

11.2 Thermometers

11.2.1 Thermometric Property (Covered in Year 4 / O-Level Physics)

All thermometers are based on some material's thermometric property. A thermometric property is a physical quantity which varies continuously with temperature in a quantitative and repeatable manner. See **Appendix A** to review necessary and desirable features in a thermometric property.

Types of Thermometer ¹	Thermometric Property
Liquid-in-glass thermometer	Length of liquid in a capillary
Thermocouple thermometer	E.m.f. between junctions of a wire at different temperatures
Thermistor	Resistance of a semiconductor
Platinum resistance thermometer	Resistance of a coil of wire
Constant-volume gas thermometer	Pressure of a fixed mass of gas at constant volume

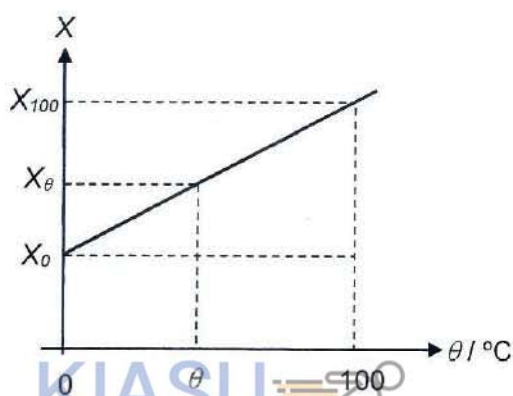
11.2.2 Centigrade Scale (Empirical scale)

The earliest scales were called **empirical scales** because they were based on experimental measurements of specific thermometric properties.

For example, the centigrade scale uses 0 °C to represent freezing point of water and 100 °C to represent boiling point of water. The temperatures between these two fixed points are divided into 100 equal intervals. Hence the centigrade temperature θ is given by the formula

$$\theta / ^\circ\text{C} = \frac{X_\theta - X_0}{X_{100} - X_0} \times 100, \text{ where}$$

X represents the thermometric property.



The **arbitrary nature of the centigrade scale** is highlighted by the fact that thermometers **based on different thermometric properties** are **only calibrated to agree at 0°C and 100°C**, and may not necessarily agree at any other intermediate temperature. (Refer to **Appendix C** for details on limitations of the centigrade scale.)

¹ For more details of thermometers and their characteristics, refer to **Appendix B**.

Worked Example 11.2.1

Say you are given a mercury-in-glass and a platinum resistance thermometer. The resistance values of the resistance thermometer are $40\ \Omega$ and $80\ \Omega$ at the ice and steam points, respectively. When both thermometers are used to measure the temperature of a cup of coffee, the mercury-in-glass thermometer reads $50\ ^\circ\text{C}$ and the resistance value of the resistance is $55\ \Omega$.

What temperature does the resistance thermometer read?

Solution:

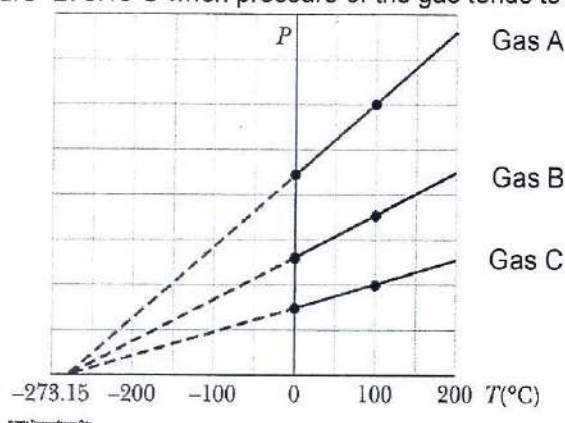
Assumption: The resistance thermometer will be used to create a centigrade scale.

	Mercury-in-glass thermometer	Resistance of Platinum	Calibrated platinum resistance thermometer
Ice point	$0\ ^\circ\text{C}$	$40\ \Omega$	$0\ ^\circ\text{C}$ (fixed point)
Steam point	$100\ ^\circ\text{C}$	$80\ \Omega$	$100\ ^\circ\text{C}$ (fixed point)
Temperature of coffee	$50\ ^\circ\text{C}$	$55\ \Omega$	Assuming that the resistance increases linearly with temperature, an increase of temperature of $100\ ^\circ\text{C}$ corresponds to an increase in resistance of $80 - 40 = 40\ \Omega$. So each increase in $1\ \Omega$ resistance corresponds to an increase in temperature of $\frac{100}{40} = 2.5\ ^\circ\text{C}$. Since the temperature of coffee corresponds to the resistance of $55 - 40 = 15\ \Omega$ above resistance of platinum at ice point, the temperature is $15 \times 2.5 = 37.5\ ^\circ\text{C}$ In short we are using this equation: $\theta / ^\circ\text{C} = \frac{X_\theta - X_0}{X_{100} - X_0} \times 100$ $\theta / ^\circ\text{C} = \frac{55 - 40}{80 - 40} \times 100 = 37.5\ ^\circ\text{C}$

11.2.3 Kelvin Scale (Absolute Scale)

In the late 1600s, the study of gas behaviour under extremely low pressures yielded two remarkable discoveries:

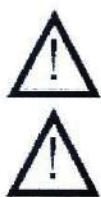
1. The pressure of a constant volume of gas varies linearly with its temperature in $^{\circ}\text{C}$; more intriguingly,
2. Regardless of the type of gas (nitrogen, oxygen, etc.), each linear curve extrapolates to the temperature -273.15°C when pressure of the gas tends to zero.



This strongly suggests that there is a lowest temperature achievable. At **absolute zero** the system has its minimum (not zero!!!) possible total internal (kinetic plus potential) energy.

Absolute zero is the lowest possible temperature. All substances have a **minimum** internal energy at this temperature.

Thus, Lord William Thomson Kelvin proposed a new temperature scale known as the ideal-gas scale, later renamed the Kelvin scale. In this scale, temperatures carry the unit of kelvin (K). This extrapolated zero-pressure temperature of -273.15°C is designated as 0 kelvin (0 K), known as **absolute zero**. The ice point and steam point correspond to 273.15 K and 373.15 K on the Kelvin scale, respectively. (Refer to **Appendix D** for details on calibration of the ideal-gas scale.)



In SI nomenclature, "degree" is not used. So it is 0 K, not 0°K . Another thing, it is always capitalised when abbreviated (K), but never capitalised when spelled (kelvin). Don't argue. He was Lord Kelvin.

Because of quantum effects, it is not true that at 0 K, all molecular motion ceases. There are theoretical reasons for believing that absolute zero cannot be attained experimentally.

11.2.4 Thermodynamic Scale (Absolute Scale)

In 1848, Lord Kelvin finally succeeded in developing an absolute thermodynamic scale based on the laws of heat transfer in an ideal heat engine. Since the ideal heat engine can be built out of any substance, this scale is truly *absolute*.

The **thermodynamic scale** is an absolute scale because it is a theoretical scale that is independent of the thermometric properties of any substance or on any empirical results.

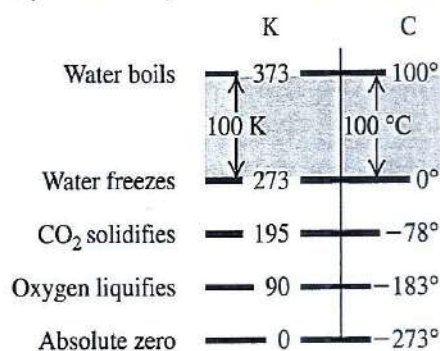
Fortunately, Lord Kelvin was also able to show that the ideal-gas scale is equivalent to the thermodynamic scale. In other words, scientists were using an absolute scale even before they knew it was an absolute scale. In celebration, the ideal-gas scale was renamed the Kelvin scale.

The Celsius Scale was later proposed. The Celsius scale is just a mere arithmetic adjustment of the thermodynamic scale measured in kelvin.

$$T / ^\circ\text{C} = T / \text{K} - 273.15$$

Notice that the units of the Kelvin scale are the same size as those on the Celsius scale. In other words, an interval of 1 K is equal to an interval of 1°C . This leads to the simple conversion between the Kelvin and Celsius scales.

The chart below illustrates some important temperatures in both the Kelvin and Celsius scales.



Wait a minute, doesn't the Celsius Scale look like the old centigrade scale? Well, the Celsius scale was designed to look like the centigrade scale, so that we can continue to call the ice point 0°C and the steam point 100°C , so that the laymen don't even notice anything has changed. However, as a student of Physics, you should understand the conceptual distinction between them. **The Celsius scale is based on the Kelvin scale, which is equivalent to the thermodynamics scale which is an absolute scale. The centigrade Scale is an empirical scale based on the thermometric property of water.**

11.3 Specific Heat Capacity

The **heat capacity** of an object is the amount of energy that is required to raise the temperature of the object by one unit of temperature.

Notes:

1. The SI unit for heat capacity is J K^{-1} .
2. The commonly used symbol for representing heat capacity is C .

The **specific heat capacity** of a substance is the energy per unit mass required to raise the temperature of the substance by one unit of temperature.

Notes:

1. The SI unit for heat capacity is $\text{J kg}^{-1} \text{K}^{-1}$.
2. The commonly used symbol for representing specific heat capacity is c .

The table below lists specific heat capacities of some substances at 25°C and 1 atmospheric pressure.

Substance	Specific heat capacity ($\text{J kg}^{-1} \text{K}^{-1}$)	Substance	Specific heat capacity ($\text{J kg}^{-1} \text{K}^{-1}$)
Water	4186	Iron	448
Ice	2090	Copper	387
Aluminium	900	Mercury	138
Glass	837	Lead	128

The amount of heat Q supplied to an object to produce a change of temperature of ΔT is given by

$$Q = C\Delta T \text{ or } Q = mc\Delta T.$$



In equations in which T appears, such as the ideal gas law, the Kelvin temperature *must* be used. In equations involving ΔT , it is possible to use either Celsius or Kelvin temperatures, because a change in temperature is the same in both scales. ($T_K \neq T_C$ but $\Delta T_K = \Delta T_C$)

Solving Problems – Revisiting Conservation of Energy

Two-Body Problems

In a typical two-body problem, two bodies initially at different temperatures exchange heat until thermal equilibrium is achieved. Assuming an isolated system (i.e. the only heat exchange is between the two bodies), the law of conservation of energy allows us to write:

Energy lost by the hotter object = Energy gained by the colder object

$$m_x c_x (T_{x,i} - T_f) = m_y c_y (T_f - T_{y,i})$$

Worked Example 11.3.1 - Thermal Energy Between Two Bodies

A 0.0500 kg metal ingot is heated to 200.0 °C and then dropped into a beaker containing 0.400 kg of water initially at 20.0 °C. If the final equilibrium temperature of the system is 22.4 °C, find the specific heat capacity of the metal. Hence, suggest what this metal could be.

Solution:

Assuming that the thermal energy absorbed by the beaker is negligible and that no energy is lost to the surroundings, we can further assume that the system of water and ingot is an isolated system.

We first look at the objects involved in the thermal transfer and look at the changes they go through:

$$\begin{array}{lcl} \text{Ingot} & : & \text{Solid} \xrightarrow{mc\Delta\theta} \text{Solid} \\ & & \theta_{\text{ingot}} = 200.0^\circ\text{C} \qquad \theta_f = 22.4^\circ\text{C} \\ \\ \text{Water} & : & \text{Liquid} \xrightarrow{mc\Delta\theta} \text{Liquid} \\ & & \theta_w = 20.0^\circ\text{C} \qquad \theta_f = 22.4^\circ\text{C} \end{array}$$

By conservation of energy,

Heat gained by the water = Heat lost by the metal

Therefore,

$$m_w c_w (\theta_f - \theta_w) = m_{\text{ingot}} c_{\text{ingot}} (\theta_{\text{ingot}} - \theta_f)$$

$$\Rightarrow (0.400 \text{ kg}) (4186 \text{ J kg}^{-1} \text{ K}^{-1}) (22.4^\circ\text{C} - 20.0^\circ\text{C}) = (0.0500 \text{ kg}) (c_{\text{ingot}}) (200.0^\circ\text{C} - 22.4^\circ\text{C})$$

$$\Rightarrow c_{\text{ingot}} = 453 \text{ J kg}^{-1} \text{ K}^{-1}$$

From the table, we see that the ingot is most likely to be iron.

Three or More Bodies Problems

If the problem involves $n > 2$ bodies exchanging thermal energy, each initially at a different temperature, it is easier to form our equation in the following manner:

$$\sum_{i=1}^n Q_i = 0$$

$$m_1 c_1 (T_{1,i} - T_f) + m_2 c_2 (T_{2,i} - T_f) + \dots m_n c_n (T_{n,i} - T_f) = 0$$

Worked Example 11.3.2 – Thermal Energy Changes Between 3 Bodies

Suppose 0.400 kg of water initially at 40.0 °C is poured into a 0.300 kg glass beaker having a temperature of 25.0 °C. A 0.500 kg block of aluminium at 37 °C is placed in the water and the system is insulated. Calculate the final equilibrium temperature of the system.

Solution:

The three objects involved in this process of heat transfer are

Water	:	$\theta_w = 40.0\text{ }^{\circ}\text{C}$	$\xrightarrow{Q_w}$	θ_f
Aluminium	:	$\theta_{al} = 37.0\text{ }^{\circ}\text{C}$	$\xrightarrow{Q_{al}}$	θ_f
Glass Beaker	:	$\theta_g = 25.0\text{ }^{\circ}\text{C}$	$\xrightarrow{Q_g}$	θ_f

When the three objects are in thermal equilibrium finally, they should be at the final temperature, θ_f . By conservation of energy, since the three objects form an isolated system, the net heat gain by the whole system is 0:

$$Q_w + Q_{al} + Q_g = 0 \quad \Rightarrow \quad m_w c_w (\theta_f - \theta_w) + m_{al} c_{al} (\theta_f - \theta_{al}) + m_g c_g (\theta_f - \theta_g) = 0$$

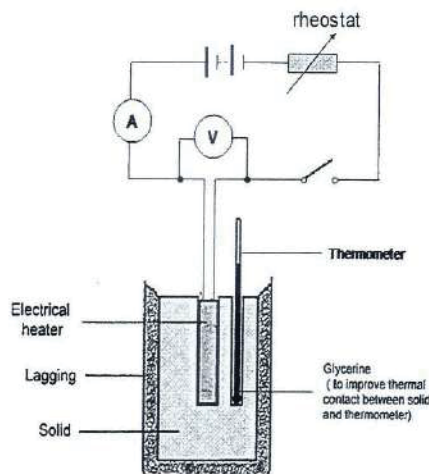
Simplifying, we have

$$\begin{aligned} \theta_f &= \frac{m_w c_w \theta_w + m_{al} c_{al} \theta_{al} + m_g c_g \theta_g}{m_w c_w + m_{al} c_{al} + m_g c_g} \\ &= \frac{(0.400)(4186)(40.0) + (0.500)(900)(37.0) + (0.300)(837)(25.0)}{(0.400)(4186) + (0.500)(900) + (0.300)(837)} \\ &= 37.8\text{ }^{\circ}\text{C} \end{aligned}$$

11.3.1 Determination of Specific Heat Capacity - Electrical Methods

Experiment 1: To determine the Specific Heat Capacity of a Solid using an Electric Heater

Setup:



Description:

Two holes are drilled into the cylindrical solid: one for the electrical heater and the other for the thermometer.

Close the switch. Use the rheostat to maintain a constant current through the heater. See that the voltmeter and ammeter readings V and I are stable.

Record the initial temperature θ_i and start the stop watch. After time interval t , record the final temperature θ_f .

Analysis:

Assuming that there are no energy losses to the surroundings,

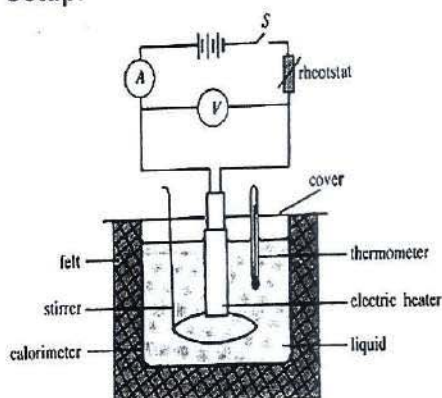
$$\begin{aligned} \text{Energy supplied by heater} &= \text{Energy gained by the solid block} \\ IVt &= mc(\theta_f - \theta_i) \end{aligned}$$

The mass of the block m can be found using a mass balance. The specific heat capacity of the solid c can then be found.

Notes:

- We keep the energy lost to the surrounding to a minimum by:
 - lagging the solid well with a good insulator of heat
 - not letting the temperature of the solid rise significantly e.g. $>10^\circ\text{C}$
 - choosing a thermometer and heater of low heat capacities so that the heat they absorb will be negligible.
- Glycerine is used to form a good thermal contact between the solid and thermometer, so that the thermometer truly measures the temperature of the solid and not the layer of air.

(The other common method that is used in determining heat capacities is known as the *method of mixtures*, which is the one which we commonly encounter in problems. This method should have been covered in the Secondary syllabus.)

Experiment 2: To determine the Specific Heat Capacity of a Liquid using an Electric Heater**Setup:****Description:**

A calorimeter of known mass m_c and specific heat capacity c_c is used to contain the liquid to be investigated.

Close the switch. Use the rheostat to maintain a constant current through the heater. See that the voltmeter and ammeter readings V and I are stable.

Record the initial temperature θ_i and start the stopwatch. After time interval t , record the final temperature θ_f .

Analysis:

Assuming that there are no energy losses to the surroundings,

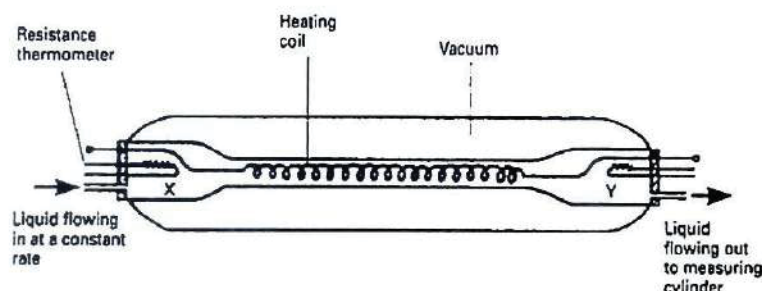
$$\begin{aligned} \text{Electrical energy supplied by the heater} \\ = \text{Energy gained by the liquid} + \text{Energy gained by the calorimeter} \\ IVt = mc(\theta_f - \theta_i) + m_c c_c (\theta_f - \theta_i) \end{aligned}$$

Notes:

- We keep the energy lost to the surroundings to a minimum by:
 - lagging the calorimeter
 - not letting the temperature of the liquid rise significantly e.g. $>10^\circ\text{C}$
 - choosing a thermometer and heater of low heat capacities so that the heat they absorb will be negligible.
- The liquid should be stirred continuously to ensure an even temperature throughout.

Note:

In general, the specific heat capacity of most solids and liquids varies slightly with temperature. For small ranges of temperature, its value may be assumed to be a constant.

Experiment 3: To determine the Specific Heat Capacity of a Liquid using Continuous Flow Method**Setup:****Fig 14.3.1c** Callendar and Barnes' apparatus. (Source: Muncaster (2000), p.241)**Description:**

The liquid to be investigated is passed through a continuous flow calorimeter at a constant mass flow rate. The liquid enters the calorimeter at temperature θ_{in} , gets heated up by the heating coil, and leaves at temperature θ_{out} .

The liquid is allowed to flow for some time, until a steady state is obtained. This happens when the temperature readings at the inlet and outlet are constant. In this steady state, the temperatures θ_{in} and θ_{out} are recorded and so are the current I through the electrical heater and the potential difference V across the heater.

The mass rate of flow of liquid m is determined by collecting and weighing the mass of liquid M from the outlet in a measured time t using $m = M/t$.

The whole experiment is then repeated with a new mass flow rate m' , but with the current and potential difference adjusted to I' and V' such that the temperatures at the outlet and inlet, θ_{out} and θ_{in} , remain the same as before.

Analysis:

Electrical energy supplied by the electrical heater
= Energy gained by the liquid + Heat Lost to the surrounding

$$\Rightarrow IVt = Mc(\theta_{out} - \theta_{in}) + Q_{surroundings}$$

Dividing the equation throughout by t ,

$$\Rightarrow IV = mc(\theta_{out} - \theta_{in}) + q \quad (1)$$

where q is the rate of heat loss to the surroundings.

Likewise, for the second set of data, we can write

$$\Rightarrow I'V' = m'c(\theta_{out} - \theta_{in}) + q \quad (2)$$

$$(2) - (1): I'V' - IV = (m' - m)c(\theta_{out} - \theta_{in}) \quad (3)$$

where c can be found.

Notes:

1. The rate of heat loss to the surroundings q is dependent on the temperature of the liquid and apparatus. Obviously, the hotter they are, the faster they lose heat. The two sets of data are collected at the same inlet and outlet temperatures so that q is the same in equations (1) and (2), allowing **heat loss to the surroundings to be eliminated through calculations** in equation (3).
2. During steady state, the temperature of the calorimeter is constant. No heat is used to heat up the calorimeter. It is thus not necessary to know the specific heat capacity of the apparatus.

3. Since the temperature readings are constant, we can use a platinum resistance thermometer, which is more accurate than a mercury thermometer, but takes more time to read.
4. This experiment can also be modified to determine the specific heat capacity of a gas by passing a gas at a constant flow rate through the continuous flow calorimeter.

Worked Example 11.3.3

In an experiment to determine the specific heat capacity of a liquid by the continuous flow method, the following two sets of data were obtained at steady-state conditions:

	1 st Set	2 nd Set
Ammeter Reading, I/A	4.3	2.6
Voltmeter Reading, V/V	2.8	1.5
Temp. at the Outlet, $\theta_{out}/^{\circ}C$	34.1	34.1
Temp. at the Inlet, $\theta_{in}/^{\circ}C$	28.6	28.6
Mass of liquid collected, m/kg	0.043	0.036
Period of Expt., t/s	60	120

- (a) Find the specific heat capacity of the liquid.
- (b) Explain whether it would have been possible to determine the specific heat capacity of the liquid if the second set of data had different inlet and outlet temperatures.

Solution:

- (a) In the continuous flow method, under steady state conditions, we have
Energy supplied by the heater = Energy gained by liquid + Heat lost to the surroundings

$$IVt = mc(\theta_{out} - \theta_{in}) + ht \quad \Rightarrow \quad IV = \frac{m}{t}c(\theta_{out} - \theta_{in}) + h$$

where h is the rate at which the system loses heat to the surrounding and c is the specific heat capacity of the liquid.

There are two sets of data, we have

$$I_1 V_1 = \frac{m_1}{t_1} c (\theta_{out} - \theta_{in}) + h \quad \dots\dots\dots(1)$$

$$I_2 V_2 = \frac{m_2}{t_2} c (\theta_{out} - \theta_{in}) + h \quad \dots\dots\dots(2)$$

$$(2) - (1), I_1 V_1 - I_2 V_2 = \left(\frac{m_1}{t_1} - \frac{m_2}{t_2} \right) c (\theta_{out} - \theta_{in})$$

Hence, the specific heat capacity of the liquid is given by

$$c = \frac{I_1 V_1 - I_2 V_2}{\left(\frac{m_1}{t_1} - \frac{m_2}{t_2} \right) (\theta_{out} - \theta_{in})} = \frac{(4.3 A)(2.8 V) - (2.6 A)(1.5 V)}{\left(\frac{0.043 \text{ kg}}{60 \text{ s}} - \frac{0.036 \text{ kg}}{120 \text{ s}} \right) (34.1^{\circ}C - 28.6^{\circ}C)}$$

$$= 3552 \text{ J kg}^{-1} \text{ K}^{-1}$$

- (b) No, it would not have been possible to find the specific heat capacity of the liquid if the second data set did not have the same outlet and inlet temperatures. This is because, if the temperatures were different, then the rate at which heat is lost to the surrounding would be different from the first case and hence the rate at which heat is lost to the surrounding cannot be eliminated from the equations in (1) and (2).

11.4 Specific Latent Heat

There are three phases of matter: solid, liquid and gas². A phase change involves a change in the internal energy, but no change in the temperature of the substance, even though heat is transferred between the substance and its surrounding.

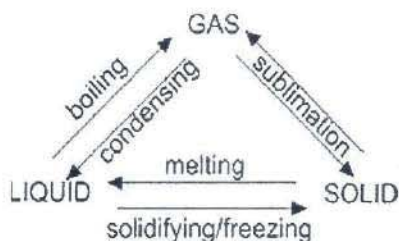


Fig. 11.4a Different processes involving phase changes

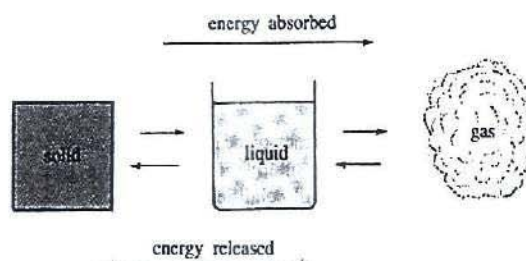


Fig. 11.4b Energy is absorbed or released during phase change.

The **specific latent heat of fusion** (L_f) is defined as the energy per unit mass required to change a substance from solid to liquid without a change in temperature.

SI unit: J kg^{-1}

The heat supplied to a system of mass m to change it from solid phase to liquid phase, without causing a change in the temperature, is thus

$$Q = mL_f$$

The **specific latent heat of vaporisation** (L_v) is defined as the energy per unit mass required to change a substance from liquid to gas without a change in temperature.

SI unit: J kg^{-1}

The heat supplied to a system of mass m to change it from liquid to gaseous phase, without causing a change in the temperature, is thus

$$Q = mL_v$$

² There is a fourth state of matter known as **plasma**, which occurs at extremely high temperatures consisting of atomic nuclei and electrons all moving at very high speeds. It has been estimated that most of the matter in the Universe (~ 90%) is actually plasma.

Worked Example 11.4.1 – Incomplete Phase Change

A 5.00 kg block of ice at 0 °C is added to an insulated container partially filled with 10.0 kg of water at 15.0 °C.

- (a) Find the final temperature, neglecting the heat capacity of the container.
 (b) Find the mass of the ice that has melted.

Strategy:

This question is tricky because the ice does not melt entirely in this example. When there is any doubt concerning whether there will be a complete phase change, it is always good to make some preliminary calculations to check.

Solution:

- (a) First, let us find the energy required to melt the ice completely:

$$Q_{\text{melt}} = m_{\text{ice}} L_f = (5.00 \text{ kg}) (3.33 \times 10^5 \text{ J kg}^{-1}) = 1.67 \times 10^6 \text{ J}$$

The maximum energy that can be lost by the initial mass of water without freezing it:

$$\begin{aligned} Q_{\text{water}} &= m_{\text{water}} c_{\text{water}} (\theta_{\text{initial}} - \theta_{\text{final}}) \\ &= (10.0 \text{ kg}) (4186 \text{ J kg}^{-1} \text{ } ^\circ\text{C}^{-1}) (15 \text{ } ^\circ\text{C} - 0 \text{ } ^\circ\text{C}) \\ &= 6.29 \times 10^5 \text{ J} \end{aligned}$$

Since the energy that can be lost by the water is not sufficient to melt all the ice, the final state of the system is a mixture of water and ice and final temperature of the system is $\theta_f = 0 \text{ } ^\circ\text{C}$.

- (b) To find the mass of ice that melted.

Since the mass of ice is already initially at melting point, the heat it gains from the water just goes to melting part of it.

Let m be the mass of the ice that has melted,

$$\begin{aligned} m L_f &= Q_{\text{water}} \\ m &= \frac{Q_{\text{water}}}{L_f} = \frac{6.29 \times 10^5 \text{ J}}{3.33 \times 10^5 \text{ J kg}^{-1}} = 1.89 \text{ kg} \end{aligned}$$

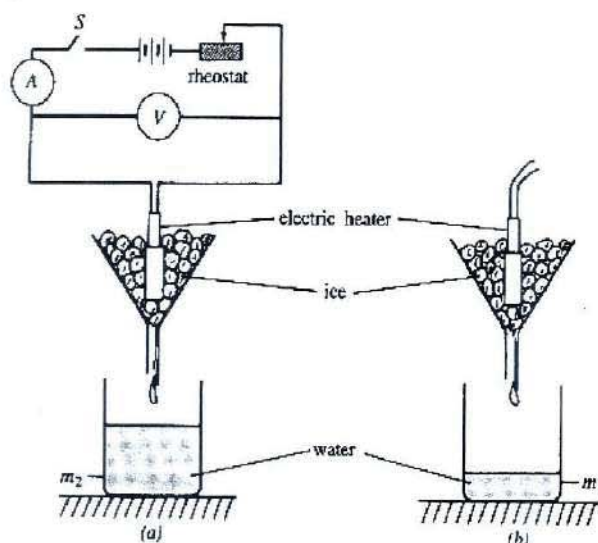
Remarks

If this problem is solved assuming (wrongly) that all the ice melts, a final temperature $\theta_f = -16.5 \text{ } ^\circ\text{C}$ would have been obtained. This result would have been nonsensical!

11.4.1 Determination of Specific Latent Heats – Electrical Methods

Experiment 1: To determine the Specific Latent Heat of Fusion of Ice using an Electric Heater

Setup:

**Fig 11.4.1a** Experimental setup to determine the specific latent heat of fusion of ice.

(a) The setup with the switch closed. (b) The setup with the switch opened as a control experiment

Description:

An electric heater is placed in a funnel containing ice. At the instant the switch S is closed, a beaker is placed below the funnel to collect the water from the melting ice. The rheostat is adjusted when necessary to maintain a constant current through the heater. After a measured time t , the beaker is removed from below the funnel and the mass m_2 of water collected is measured.

To take into consideration the mass of ice melted due to absorption of heat from the surroundings, the same amount of ice is placed in another funnel containing an electric heater which is not connected to the circuit. The mass m_1 of ice melted after the measured time t is obtained by collecting the water with a beaker.

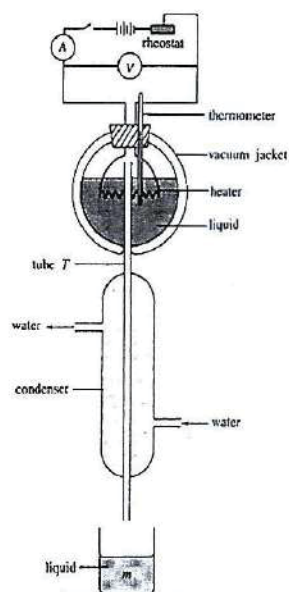
Analysis:

Mass of ice melted by electric heater = $m_2 - m_1$

By conservation of energy,

$$\begin{aligned} \text{Energy supplied by the electrical heater} &= \text{Energy used to melt the ice} \\ IVt &= \text{of mass } (m_2 - m_1) \\ &= (m_2 - m_1) L_f \end{aligned}$$



Experiment 2: To determine the Specific Latent Heat of Vaporisation using an Electric Heater**Setup:****Description:**

In this method, the liquid to be investigated is placed in a flask which is surrounded by a vacuum jacket and is heated by an electrical heater. The liquid is heated until it starts to boil.

After a measured time t , the mass m of the liquid that has vaporised is measured by collecting the liquid condensed by the condenser using a beaker.

The whole experiment is then repeated with a different setting of the rheostat for a different rate of heat supply.

Analysis:

By conservation of energy,

Energy supplied by the electrical heater =

Energy required to vaporize the liquid + Total amount of heat lost to surrounding

$$\Rightarrow IVt = mL_v + ht$$

where h is the rate at which heat is lost to the surroundings

Dividing through by t , we have $IV = \frac{m}{t}L_v + h$

Since we have two sets of data, we can write two equations:

$$\text{Trial 1 : } I_1 V_1 = \frac{m_1}{t_1} L_v + h \quad (1)$$

$$\text{Trial 2 : } I_2 V_2 = \frac{m_2}{t_2} L_v + h \quad (2)$$

$$(1) - (2): I_1 V_1 - I_2 V_2 = \left(\frac{m_1}{t_1} - \frac{m_2}{t_2} \right) L_v \quad (3)$$

Hence, L_v may be found.

Notes:

1. When the liquid is boiling, the temperature of the liquid remains constant. Hence, the rate at which heat is lost to the surroundings becomes constant. This allows **heat loss to the surroundings to be eliminated through calculations** in equation (3).
2. Furthermore, since the temperature of the apparatus no longer rises, no energy is required to raise its temperature and hence the specific heat capacity of the experimental setup need not be considered.

Worked Example 11.4.2 – Specific Latent Heat of Vaporisation of a Liquid

In an electrical method to determine the specific latent heat of vaporisation of a liquid, the liquid was heated by means of a heating coil and the steady-state quantities measured under two different operating conditions are given in the table below.

	Trial 1	Trial 2
Current in heating coil	2.5 A	3.0 A
Potential difference across coil	15.0 V	18.0 V
Mass of liquid vaporised in $t = 250$ s	38.5 g	59.6 g

- (a) Calculate the specific latent heat of vaporisation of the liquid.
 (b) Estimate the rate of loss of heat to the surroundings.

Solution:

- (a) By conservation of energy,

for trial 1

$$\begin{array}{rclcl}
 \text{Power} & & \text{Rate at which} & & \text{Rate at which} \\
 \text{supplied by the} & = & \text{the energy is} & + & \text{heat is lost to the} \\
 \text{heater} & & \text{supplied to} & & \text{surroundings} \\
 & & \text{vaporize the} & & \\
 & & \text{liquid} & & \\
 \Rightarrow IV & = & \frac{m}{t}L & + & h \quad \dots\dots\dots (1)
 \end{array}$$

Similarly, for trial 2,

$$\Rightarrow I'V' = \frac{m'}{t}L + h \quad \dots\dots\dots (2)$$

$$\begin{aligned}
 (2) - (1), \text{ we have } \quad I'V' - IV &= \left(\frac{m'}{t} - \frac{m}{t}\right)L \\
 \Rightarrow L &= \frac{(I'V' - IV)t}{(m' - m)} = \frac{(3.0 \times 18.0 - 2.5 \times 15.0)(250)}{(59.6 - 38.5) \times 10^{-3}} = 1.95 \times 10^5 \text{ J kg}^{-1}
 \end{aligned}$$

- (b) From equation (1),

$$\begin{aligned}
 \text{The rate of heat loss, } h &= IV - \frac{mL}{t} \\
 &= 3.0 \times 18.0 - \frac{(59.6 \times 10^{-3})(1.95 \times 10^5)}{250}
 \end{aligned}$$

$$= 7.51 \text{ W}$$

KIASU
ExamPaper

11.5 Kinetic Model of Matter

The kinetic theory of matter states that all matter is made up of a large number of tiny atoms or molecules which are in continuous motion.

PHASES OF MATTER	STRUCTURE AND CHARACTERISTICS	MOVEMENT OF MOLECULES - A KINETIC MODEL
Solid 	Atoms or molecules are closely packed (a) Crystalline solid — atoms or molecules are compactly arranged in a regular and orderly lattice structure. Crystals have a high density. (b) Non-crystalline solid — atoms or molecules are arranged in a more random way. Density is still high but generally lower than many crystalline solids.	Atoms or molecules are locked in and constrained by attractive and repulsive intermolecular forces. They vibrate about their equilibrium positions continually. Molecular motion is minimum at zero kelvin. As temperature rises, molecular motion and kinetic energy increase.
	Has fixed shape and volume. Solids are not easily compressed or stretched by mechanical means.	When solids are stretched, atoms or molecules move further apart but are opposed by attractive forces. When solids are compressed, atoms/molecules move closer together but are opposed by repulsive forces.
Liquid 	Atoms or molecules are only slightly further apart than in solids. Density remains high.	Atoms or molecules vibrate, but in addition have random and restricted translational motion, since they occur in clusters.
	Volume is nearly constant but has variable shape. Liquids are not easily compressed by mechanical means.	Attractive forces between atoms or molecules are weaker. This allows for greater mobility which in turn explains why liquids can take the shape of any container. Attractive forces manifest themselves as surface tension and viscosity of a liquid.
Gas 	Atoms or molecules are very far apart. Densities of gases are very low.	Atoms or molecules move randomly at very high speed throughout the space the atoms or molecules occupy. The average kinetic energy is directly proportional to temperature (in K).
	Gases do not have a fixed volume or shape. They are highly compressible.	Atoms or molecules will occupy any available space in a container. Intermolecular forces between them are negligible as they are very far apart.

11.5.1 Phase Changes at a Microscopic Level

Fig 11.5.1 is a plot of the experimental measurement of temperature as energy is added to some solid at some initial temperature below its melting point.

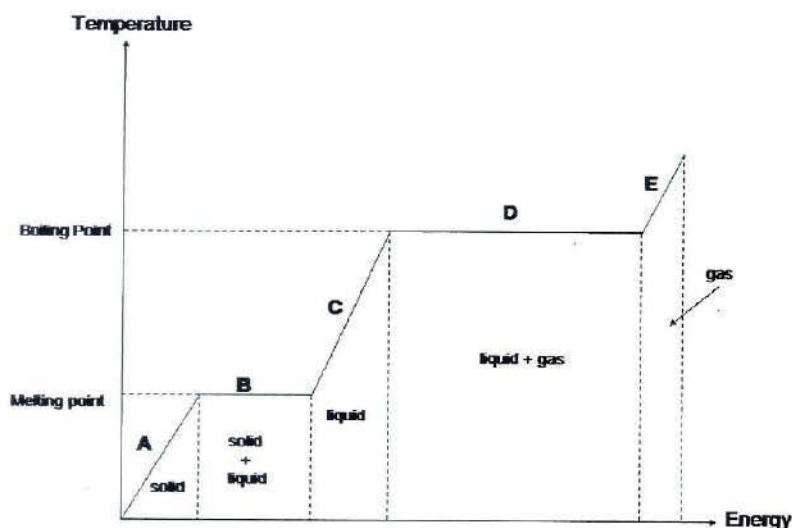


Fig 11.5.1

- **Parts A, C and E:** For these parts of the graph, we clearly see that the energy added into the system increases the temperature of the system and hence, at the molecular level, will increase the microscopic random kinetic energy of the molecules.
- **Parts B and D:** In these parts of the graph, the system remains at the same temperature. Hence, there is no increase in the microscopic random kinetic energy of the molecules despite heat being added to it. *So what happened to the energy supplied?*

11.5.1.1 Melting

The process during which a solid changes to a liquid at melting point is known as *melting* (Part B) and the energy supplied to bring about the change is known as the latent heat of fusion.

During melting, the substance is transformed from the ordered solid phase to the disordered liquid phase. Work must be done to enable the molecules to overcome the attractive bonding between them, so that they can slip out of the rigid structure to an arrangement with weaker intermolecular forces. Latent heat provides the energy for that.

Since the intermolecular force is weaker in a liquid, the potential energy is higher (less negative). The heat supplied during melting goes only into increasing the potential energy component of the internal energy. The temperature does not rise since the kinetic-energy component remains constant during melting.

11.5.1.2 Boiling

Similarly, in Part D, a process called *boiling* occurs. In this case, the system changes from the liquid phase to the gaseous phase at boiling point.

During boiling, the molecules are totally liberated, ending up in a state in which there is virtually zero intermolecular force. Work must be done to enable molecules to overcome the attractive bonding and allow them to move around independently as gas molecules. Latent heat provides the energy for that.

Since the intermolecular force is virtually zero in a gas, the potential energy is virtually zero. The heat supplied during boiling goes only into increasing the potential energy component of the internal energy. The temperature does not rise since the kinetic-energy component remains constant during boiling.

11.5.1.3 Comparison between L_f and L_v

The table below lists some of the specific latent heats of fusion and vaporisation of various substances.

Substance	Melting Point / °C	Latent Heat of Fusion / J kg ⁻¹	Boiling Point / °C	Latent Heat of Vaporisation / J kg ⁻¹
Helium	-269.65	5.23×10^3	-268.93	2.09×10^4
Nitrogen	-209.97	2.55×10^4	-195.81	2.01×10^5
Oxygen	-218.79	1.38×10^4	-182.97	2.13×10^5
Ethyl Alcohol	-114	1.04×10^5	78	8.54×10^5
Water	0.00	3.33×10^5	100.00	2.26×10^6
Sulphur	119	3.81×10^4	444.60	3.26×10^5
Lead	327.3	2.45×10^4	1 750	8.70×10^5
Aluminium	660	3.97×10^5	2 450	1.14×10^7
Silver	960.80	8.82×10^4	2 193	2.33×10^6
Gold	1 063.00	6.44×10^4	2 660	1.58×10^6
Copper	1 083	1.34×10^5	1 187	5.06×10^6

Examining the table, you can observe that the specific latent heat of vaporisation for a given substance is higher than the specific latent heat of fusion.

There are two reasons for this:

- (1) The required increase in potential energy for boiling is usually much larger than that for melting. The intermolecular bonds are completely broken down during boiling. Hence the amount of bond weakening (and thus the increase in potential energy) during boiling is much larger than that during melting.
- (2) Work must be done against the atmospheric pressure during boiling and less so during melting. During melting, the molecules remain bonded. The change in volume of the substance is insignificant and hence very little work is done against the atmosphere. During boiling, however, the liquid-type bonds are broken down completely. The expansion of volume is much more significant and hence the work done against the atmosphere is also more significant.

Therefore, it is not surprising that more energy is required to vaporize a given mass of substance than to melt it.

11.5.1.4 Evaporation

Evaporation is a process during which liquid changes phase to become a vapour. It takes place at the surface of the liquid and at any temperature.

Unlike boiling, evaporation can occur at any temperature. Boiling is in fact a special case of evaporation, where evaporation takes place at the boiling point of the liquid. At this boiling point, the rate of evaporation of the liquid is greatest.

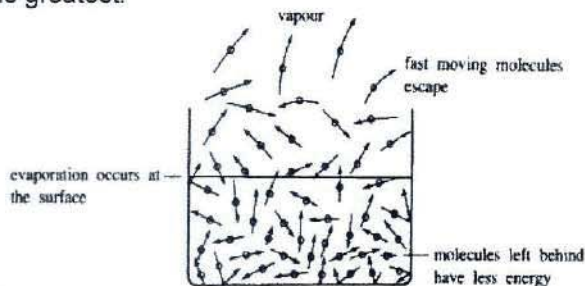


Fig 11.5.1.4 Evaporation occurring at the surface of a liquid.

The average kinetic energy per molecule is proportional to the thermodynamic temperature. For a given temperature there is a range of kinetic energies around this average. Transfer of energy takes place when molecules collide with each other. A molecule near the surface of the liquid which possesses more kinetic energy than the potential energy (energy which binds it to the rest of the liquid) is able to leave the liquid and become a molecule in the vapour state.

It follows that the rate of evaporation can be increased by the following ways:

(1) **Increase the surface area of the liquid.**

(2) **Increase the temperature of the liquid.**

An increase in temperature will increase the average kinetic energy of the molecules. Hence, there is a greater proportion of molecules with higher kinetic energies and, therefore, the probability of the molecules acquiring sufficient energy to escape at the surface is also greater.

(3) **Reducing the air pressure just above the surface liquid.**

Microscopically, it translates to having less air molecules present just above the surface of the liquid. This decreases the chance of a vapour molecule rebounding off an air molecule to return to the liquid.

(4) **Having a draught above the liquid surface.**

This will remove the vapour molecules before they have a chance to return to the liquid.

Cooling Effect of Evaporation

During the process of evaporation, the more energetic molecules escape. The molecules that remain in the liquid are less energetic and hence have a lower average kinetic energy. Macroscopically, this is manifested as a decrease in the temperature of the liquid. This explains why evaporation is often accompanied by a cooling effect.



11.6 Ideal Gases

Chemically, different gases behave very differently, e.g. some are combustible, others are inert and reactions may take place when you mix them. However, physically, all gases behave identically, e.g. for a given amount of any gas, a specific increase in pressure at constant temperature will result in a specific decrease in volume. On a microscopic scale, forces are at play between the different molecules that will affect the short-term behaviour. However, in most cases, we may ignore these microscopic effects. We can then treat the gas as a so-called ideal gas. Air at sea level and room temperature, the atmospheres of stars like our sun and most of the gases in laboratories on earth all behave like ideal gases.

11.6.1 Moles and Avogadro's Constant

The amount of gas is most often expressed in terms of the number of moles, n .

One **mole** of any substance is that amount of substance that contains the Avogadro's number $N_A = 6.02 \times 10^{23}$ of constituent particles (atoms or molecules).

The **Avogadro constant** is defined as the number of atoms in 0.012 kg of carbon-12.

The molar mass of a substance is defined as the mass of that substance divided by the number of moles of the substance. The number of moles n of a substance is related to its mass m and molar mass M

through the expression $n = \frac{m}{M}$.

Numerically, molar mass is simply the atomic mass expressed in g mol^{-1} .

For example, the molar mass of helium gas (consisting of He atoms) is $4.00 \times 10^{-3} \text{ kg mol}^{-1}$. The molar mass of oxygen gas (consisting of O_2 molecules) is $16.0 \times 2 = 32.0 \times 10^{-3} \text{ kg mol}^{-1}$. Molar masses are also commonly given in g mol^{-1} , e.g. the molar mass of oxygen gas is 32.0 g mol^{-1} .

The molar volume of a gas is the volume occupied by one mole of the gas. The molar volume V_m is equal to the molar mass divided by the density. The unit is $\text{m}^3 \text{ mol}^{-1}$.

11.6.2 The Ideal Gas Equation

An ideal gas is a theoretical gas composed of a set of randomly moving, non-interacting point particles. The ideal-gas concept is useful because it obeys the *ideal gas law*, a simplified equation of state, and is amenable to analysis under statistical mechanics. You will have learnt about the ideal-gas law in Chemistry.

Early experiments in the 17th and 18th century with ideal gases resulted in three laws known as the gas laws. They are, in chronological order, Boyle's Law, Charles' Law and the Pressure Law. The observations made in these laws were summarized in the **equation of state for an ideal gas** which is known as the **Ideal Gas Law**³.

Ideal Gas Law (in terms of no. of moles) :

$$PV = nRT \quad [\text{Need to be able to recall and use as stated in LO}]$$

where P : pressure of the gas
 V : volume occupied by the gas
 n : amount of gas in terms of number of moles
 T : thermodynamic temperature of the gas (in kelvins)
 R : Universal Gas constant

Ideal Gas Law (in terms of no. of molecules) :

$$PV = NkT$$

where N : amount of gas in terms of number of gas molecules
 k : Boltzmann's constant

Notes:

1. Temperatures used in gas laws MUST ALWAYS be in kelvin.
2. The **Universal Gas Constant** R is a constant that is experimentally determined. In SI units,

$$R = 8.31 \text{ J mol}^{-1} \text{ K}^{-1}$$

3. The **Boltzmann constant** is related to R by the relation $k = \frac{R}{N_A}$ and has a value of

$$k = \frac{R}{N_A} = 1.38 \times 10^{-23} \text{ J K}^{-1}$$



Be very clear that the k here is NOT the spring constant, it is the Boltzmann constant k which is another constant.

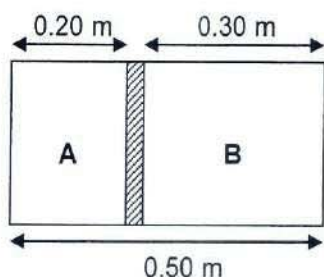
We can now formally define an ideal gas in terms of its macroscopic quantities.

An **ideal gas** is a gas which obeys the *Equation of State for the Ideal Gas* at all values of pressure, volume and temperature.

Note:

An ideal gas does not exist. Nevertheless, real gases like hydrogen, nitrogen, oxygen and carbon dioxide behave almost identically at normal pressures and temperatures. The ideal gas model tends to fail at very low temperatures or high pressures, when intermolecular forces and molecular size become important. It also fails for most heavy gases, such as water vapour and many refrigerants. At some point of low temperature and high pressure, real gases may turn into a liquid or a solid. The model of an ideal gas, however, does not allow such changes.

³ The evolution of how Boyle's, Charles' and the Pressure Law became the Ideal Gas Law is laid out in Appendix E.

Example 11.6.1 – Ideal Gas Law Application

A cylinder, 0.50 m long and of cross-sectional area S , closed at each end, is fitted with a smooth thermally insulating piston, so as to contain gases in each end **A** and **B**.

Initially, the piston is in equilibrium at a distance of 0.20 m from one end of the cylinder, and the gas in each compartment is at a temperature of 27°C .

The temperature of the smaller volume of gas, **A**, is now increased to 177°C , while that of the gas in the larger volume **B** is maintained at 27°C . The piston moves through a distance x until a new position of equilibrium is obtained.

- State the equation relating the initial and final pressures, volumes and temperatures of the gas in **A**.
- State the equation relating the initial and final pressures, volumes and temperatures of the gas in **B**.
- Hence find the new position of equilibrium.

Solution:

a) Before p_i , $V_i = 0.20S$, $T_i = 27^\circ\text{C} = 300.15\text{K}$

After p_f , $V_f = (0.20 + x)S$, $T_f = 177^\circ\text{C} = 450.15\text{K}$

Using

$$\frac{p_i V_i}{T_i} = \frac{p_f V_f}{T_f}$$

we obtain $\frac{p_i (0.20S)}{300.15} = \frac{p_f (0.20 + x)S}{450.15}$ — (1)

b) Before p_i , $V_i = 0.30S$, $T_i = 27^\circ\text{C} = 300.15\text{K}$

After p_f , $V_f = (0.30 - x)S$, $T_f = 27^\circ\text{C} = 300.15\text{K}$

Using

$$p_i V_i = p_f V_f \quad \text{as } T_i = T_f$$

we obtain $p_i (0.30S) = p_f (0.30 - x)S$ — (2)

c) (1)/(2): $\frac{(0.20S)}{300} \times \frac{1}{0.30S} = \frac{(0.20 + x)S}{450} \times \frac{1}{(0.30 - x)S}$

$$\therefore 0.20 + x = 0.30 - x$$

$$x = 0.05\text{m}$$

The new position is 0.25 m from the left.

11.6.3 Molecular Interpretation of Pressure

Among other things, the **Kinetic Theory of Gases** assumes that a gas consists of a large number of molecules in rapid random motion, colliding elastically with one another and the walls of the container.

11.6.3.1 Microscopic Model of an Ideal Gas

The kinetic theory of gases consists of a set of basic assumptions for the properties of gas molecules of an ideal gas. The assumptions are

- (A1) A gas consists of a very large number of molecules.
- (A2) The gas molecules are moving randomly.
- (A3) The collisions of the gas molecules with one another and with the walls of the container are perfectly elastic.
- (A4) There are no intermolecular forces of attraction.
- (A5) The volume of the gas molecules is negligible as compared to the volume of the container, which is the volume of the gas.
- (A6) The duration of collisions is negligible compared to the time interval between collisions.

This microscopic model of an ideal gas leads to the macroscopic description using the equation of state, $pV = nRT$.

When the gas molecules collide elastically with the container of the gas, they rebound off the wall with the same speed. Clearly, the walls of the container exert a force on the gas molecules to cause their momentum to change in direction. (By Newton's 2nd Law, this force is equal to the rate of change of momentum of the molecule.) By Newton's 3rd Law, the molecules exert an equal but opposite force on the walls.

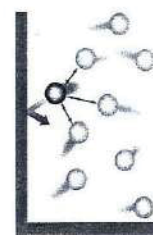


Fig. 11.6.3a

The **pressure by the gas on the container** is the **average force per unit area exerted by the molecules on the walls of the container**.

The pressure P due to the gas molecules on a wall of the container can be written as

$$\begin{aligned}
 P &= \frac{\text{average force exerted on the wall by molecules}}{\text{area of the wall}} \\
 &= \frac{(\text{no. of collisions per unit time})(\text{change of momentum per collision})}{\text{area of the wall}}
 \end{aligned}$$

We can now explain how the following factors can lead to an increase in pressure of a gas:

- **No. of molecules in the container.** All other conditions being constant, a larger number of molecules will imply more collisions made with the wall of the container.
- **Volume of container.** Having a smaller container implies that the molecules need to travel less distance to reach the next wall after rebounding from one. The no. of collisions they make with the walls per unit time therefore increases. (Refer to Fig 11.6.3b.)
- **Mass of the molecule.** The type of gas used will determine the mass of the molecule which in turn determines the momentum that each molecule possesses. The larger the mass, the greater the momentum and thus the change of momentum of the molecules per collision is greater.
- **Velocities of the molecules.** The momentum of each gas molecule is also determined by its velocity. A molecule possessing a larger velocity will have a greater momentum and hence its change in momentum will correspondingly be larger per collision.

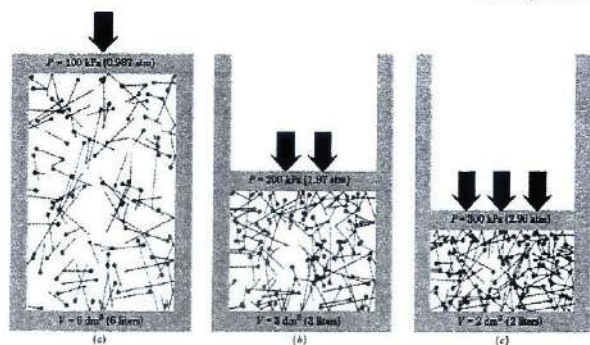


Fig. 11.6.3b There are equal numbers of molecules in the enclosed portion of each diagram. As the volume of the gas decreases, the number of molecules per unit volume increases. The number of molecules striking a given area of the walls per unit time also increases, thus increasing the force on that area. This force per unit area is the pressure and hence the pressure increases.

11.6.3.2 Relationship between the Pressure of a gas in terms of its Microscopic Quantities

From the Kinetic Theory of Gases, a gas consists of many molecules in constant random motion, colliding elastically with each other and the walls of the container. The pressure of the gas is due to the collision of the gas molecules with the walls of the container.

We shall attempt to relate the macroscopic quantity, pressure, with the microscopic model of gases as a collection of atoms moving in random motion in a vessel. [Need to know the derivation of the

relationship $pV = \frac{1}{3} Nm \langle c^2 \rangle$ as stated in LO]

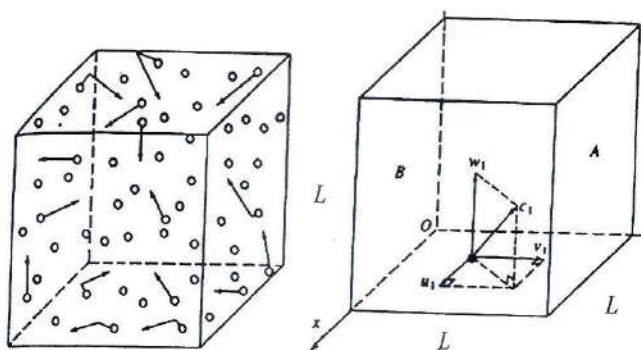


Fig.F1.

Step 1

Resolve the motion of molecule along x-y-z directions

Consider a gas of N molecules in a cubical container of sides L (Fig. F1). Suppose the mass of a molecule is m . Consider a molecule moving with a velocity c_1 . The velocity c_1 of the molecule can be resolved into x , y and z directions as u_1 , v_1 and w_1 respectively.

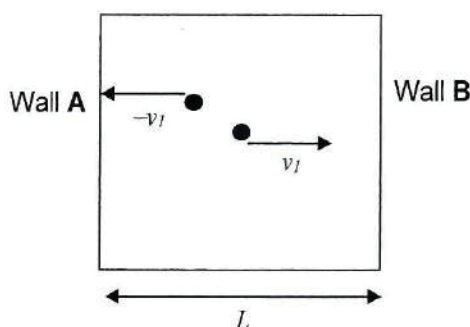
Using Pythagoras theorem, $c_1^2 = u_1^2 + v_1^2 + w_1^2$.

Step 2

Find the change of momentum upon collision with a wall

Consider the component of velocity $-v_1$ along the y -axis. After colliding with the wall A, the molecule rebounds with a velocity v_1 since the collision is perfectly elastic.

$$\begin{aligned} \text{Momentum before collision} &= -mv_1 \\ \text{Momentum after collision} &= mv_1 \\ \therefore \text{change of momentum} &= mv_1 - (-mv_1) \\ &= 2mv_1 \end{aligned}$$



Step 3**Find the force on the wall due to the bombardments of molecules on the wall**

Assuming that the molecule does not collide with other molecules, the time interval t between a collision of the molecule with the wall A and its next collision with the same wall is the same as the time the molecule takes to move from wall A to wall B and back and is given by:

$$t = \frac{2L}{v_1}$$

Hence, the number of collision with the wall A per second = $t^{-1} = \frac{v_1}{2L}$

Since the force exerted on the wall by this molecule = rate of change of momentum

$$= 2mv_1 \times \frac{v_1}{2L} = \frac{mv_1^2}{L}$$

If there are N molecules, each with the component of velocity along the y -axis as $v_1, v_2, v_3, \dots, v_N$ respectively, then the total force on wall A due to the collision of these N molecules is

$$F = \frac{mv_1^2}{L} + \frac{mv_2^2}{L} + \frac{mv_3^2}{L} + \dots + \frac{mv_N^2}{L}$$

Step 4**Find the pressure on the wall**

Pressure on wall A,

$$p = \frac{F}{L^2} = \frac{\frac{mv_1^2}{L} + \frac{mv_2^2}{L} + \frac{mv_3^2}{L} + \dots + \frac{mv_N^2}{L}}{L^2} = \frac{m}{L^3} (v_1^2 + v_2^2 + v_3^2 + \dots + v_N^2)$$

where L^3 = volume of container.

Step 5**Express in terms of root-mean-square speed v_{rms}**

Since $c_1^2 = u_1^2 + v_1^2 + w_1^2$

$$c_2^2 = u_2^2 + v_2^2 + w_2^2$$

⋮

$$c_N^2 = u_N^2 + v_N^2 + w_N^2$$

Adding all equations and dividing by N , we have

$$\frac{c_1^2 + c_2^2 + c_3^2 + \dots + c_N^2}{N} = \frac{u_1^2 + u_2^2 + u_3^2 + \dots + u_N^2}{N} + \frac{v_1^2 + v_2^2 + v_3^2 + \dots + v_N^2}{N} + \frac{w_1^2 + w_2^2 + w_3^2 + \dots + w_N^2}{N}$$

$$\langle c^2 \rangle = \langle u^2 \rangle + \langle v^2 \rangle + \langle w^2 \rangle$$

Since the motion of the gas molecules are random in nature and the number of molecules is large, on average,

$$\langle u^2 \rangle = \langle v^2 \rangle = \langle w^2 \rangle$$

This implies

$$\langle c^2 \rangle = \langle v^2 \rangle + \langle v^2 \rangle + \langle v^2 \rangle$$

$$\langle v^2 \rangle = \frac{1}{3} \langle c^2 \rangle$$

$$p = \frac{Nm}{V} \times \frac{1}{N} (v_1^2 + v_2^2 + v_3^2 + \dots + v_N^2) = \frac{Nm}{V} \times \frac{1}{3} \langle c^2 \rangle = \frac{1}{3} \frac{Nm}{V} \langle c^2 \rangle$$

And rearranging, we have

$$pV = \frac{1}{3} Nm \langle c^2 \rangle \quad \text{.....Equation \#}$$

where N = no. of molecules, m = mass of 1 molecule (also known as molecular mass) and $\langle c^2 \rangle$ is the mean square speed of the molecules.

Notes:

1. There are a few ways to rewrite Equation #.

- Since N is the total number of molecules and m is the mass of each molecule, then

$$\Rightarrow \text{Total mass of the gas} = Nm$$

If the **molar mass of the gas** is M and there are n moles of gas in the container,

$$\Rightarrow \text{Total mass of the gas} = Nm = nM$$

and Equation # can be written as

$$pV = \frac{1}{3} nM \langle c^2 \rangle \quad \text{(No need to memorise)}$$

- We can also write Equation # as

$$\Rightarrow p = \frac{1}{3} \frac{Nm \langle c^2 \rangle}{V} = \frac{1}{3} \frac{(\text{Mass of gas})}{(\text{Volume of gas})} \langle c^2 \rangle$$

$$\Rightarrow p = \frac{1}{3} \rho \langle c^2 \rangle \quad \text{(No need to memorise)}$$

where ρ is the density of the gas.

11.6.4 Molecular Interpretation of Temperature

The motion of molecules in a gas is completely random. As such, they collide with each other at irregular time intervals and over a range of angles. In these collisions, some molecules will gain kinetic energy while others will lose it. As a result, at any instant, the molecules have a range of speeds and kinetic energies.

It turns out that the **mean kinetic energy of a molecule of an ideal gas is proportional to the thermodynamic temperature.**

$$\langle KE \rangle \propto T$$

For monatomic gases, whose atoms are only capable of translational motion, and diatomic and polyatomic gases whose molecules have negligible vibrational and rotational motion, the constant of proportionality is $\frac{3}{2}k$, with Boltzmann's constant k as above; $k = 1.38 \times 10^{-23} \text{ J K}^{-1}$.

$$\langle KE \rangle = \frac{3}{2} kT$$

For a gas consisting of N molecules of molecular mass m , the average KE can be calculated by

$$\langle KE \rangle = \frac{\frac{1}{2}mv_1^2 + \frac{1}{2}mv_2^2 + \dots + \frac{1}{2}mv_N^2}{N}, \text{ where } v_n \text{ is the speed of the } n\text{th molecule.}$$

$$\text{which leads us to } \langle KE \rangle = \frac{1}{2}m \left(\frac{v_1^2 + v_2^2 + \dots + v_N^2}{N} \right)$$

We thus introduce two useful quantities called the mean-square speed $\langle v^2 \rangle$, and root-mean-square speed v_{rms} of the molecules.

$$\langle v^2 \rangle = \frac{v_1^2 + v_2^2 + \dots + v_N^2}{N}$$

$$v_{rms} = \sqrt{\langle v^2 \rangle} = \sqrt{\frac{v_1^2 + v_2^2 + \dots + v_N^2}{N}}$$

$$\text{So } \langle KE \rangle = \frac{1}{2}m \langle v^2 \rangle = \frac{1}{2}m v_{rms}^2$$

The root-mean-square velocity v_{rms} is introduced here, as v_{rms}^2 is proportional to the average kinetic energy of the individual molecules. The average velocity itself, however, is equal to zero.

Putting everything together we have

$$\frac{1}{2}m\langle v^2 \rangle = \frac{3}{2}kT$$

[Need to know this relationship as stated in LO]

Note:

1. The average translational kinetic energy of a molecule in an ideal gas depends only on the temperature, not on the pressure or the type of gas.
2. The **temperature** T of a gas is a **measure of the average translational kinetic energy** of a molecule in the gas.
3. Temperature is a large scale manifestation of the motion of the molecules. An increase in temperature corresponds to an increase in the average translational kinetic energies of the molecules.

Example 11.6.2 – Distinguishing various “speeds”

The velocities of particles moving along the x-axis are as follows:

Velocity v / ms^{-1}	-3.0	-2.0	-1.0	0.0	1.0	2.0	3.0
No. of particles	1	3	4	3	2	1	3

- What is
- (a) the average velocity of the particles?
 - (b) the average speed of the particles?
 - (c) root-mean-square speed of the particles?

Solution:

$$a) \frac{(-3.0) + 3(-2.0) + 4(-1.0) + 3(0.0) + 2(1.0) + (2.0) + 3(3.0)}{1+3+4+3+2+1+3} = 0 \text{ m/s}$$

$$b) \frac{(3.0) + 3(2.0) + 4(1.0) + 3(0.0) + 2(1.0) + (2.0) + 3(3.0)}{1+3+4+3+2+1+3} = 1.53 \text{ m/s}$$

$$c) v_{\text{rms}} = \sqrt{\langle v^2 \rangle} = \sqrt{\frac{v_1^2 + v_2^2 + v_3^2 + \dots + v_N^2}{N}}$$

$$v_{\text{rms}} = \sqrt{\frac{1(-3.0)^2 + 3(-2.0)^2 + 4(-1.0)^2 + 3(0.0)^2 + 2(1.0)^2 + 1(2.0)^2 + 3(3.0)^2}{1+3+4+3+2+1+3}}$$

$$= 1.8 \text{ m/s}$$

11.6.5 Internal Energy of an Ideal Gas

Recall that the internal energy U of a system is given by

$$U = \sum KE_{\text{microscopic}} + \sum PE_{\text{microscopic}}$$

For a system of ideal gas, the intermolecular forces between the gas molecules in an ideal gas is assumed to be negligible. The potential energies between the molecules are thus zero. Hence, the internal energy of an ideal gas is given by

$$U = \text{Total KE of molecules}$$

Since the mean KE of gas molecules is $\frac{3}{2}kT$, the total internal energy of an ideal gas consisting of N molecules is given by

$$U = \frac{3}{2}NkT = \frac{3}{2}nRT$$

Note:

The internal energy of an ideal gas depends only on the temperature of the system and the amount of gas in the system.

Example 11.6.3 – Various Microscopic Properties of an Ideal Gas

A cylinder contains 2.00 mol of helium gas at 20.0 °C. Assume that helium behaves like an ideal gas.

(a) Find the total internal energy of the system.

(b) What is the average kinetic energy per molecule?

(c) How much energy would have to be added to the system to double the rms speed?

The molar mass of helium is $4.00 \times 10^{-3} \text{ kg mol}^{-1}$.

Solution:

$$a) U = \frac{3}{2}nRT = \frac{3}{2}(2.00)(8.31)(273.15 + 20) = 7308 \text{ J}$$

$$b) \text{Avg. KE per molecule} = \frac{7308}{2(6.02 \times 10^{23})} = 6.07 \times 10^{-21} \text{ J}$$

$$c) KE = \frac{1}{2}m(v_{\text{rms}})^2 \xrightarrow{\text{constant } m} KE \propto v_{\text{rms}}^2$$

$$\frac{KE'}{KE} = \frac{(v'_{\text{rms}})^2}{(v_{\text{rms}})^2} = \frac{(2v_{\text{rms}})^2}{(v_{\text{rms}})^2} = 4$$

The kinetic energy is 4 times larger when rms speed is doubled.

$\therefore$ 3 times more energy has to be added ($7308 \times 3 = 21924 \text{ J}$)

11.7 First law of thermodynamics

First Law Of Thermodynamics

The increase in the internal energy of a system is the sum of the heat supplied to the system and the work done on the system.

Expressed in a mathematical equation,

$$\Delta U = Q + W_{on}$$

where ΔU : INCREASE in internal energy of the system
 Q : heat supplied TO the system
 W_{on} : work done ON the system

A system may be a solid, liquid or gas. Let's now look at each term in the equation more closely, for a system of a fixed amount of gas.

11.7.1 Heat, Q

This term refers to the transfer of thermal energy through conduction, convection or radiation. Q is positive when heat flows into the gas (e.g. by bringing a flame to it), negative when heat flows out of the gas (e.g. by leaving it in a cold surrounding).

11.7.2 Work, W

Work is done whenever the volume of a gas changes, much like work is done whenever the point of application of a force is displaced.

The formulae for calculation of work done ON and BY the gas are listed below. Read on for further elaboration.

	Work done ON the gas	Work done BY the gas
Constant pressure	$W_{on} = -p(V_f - V_i)$	$W_{by} = p(V_f - V_i)$
Varying pressure	$W_{on} = - \int_{V_i}^{V_f} p \, dV$	$W_{by} = + \int_{V_i}^{V_f} p \, dV$
When gas expands	negative	positive
When gas contracts	positive	negative

11.7.2.1 Derivation of W

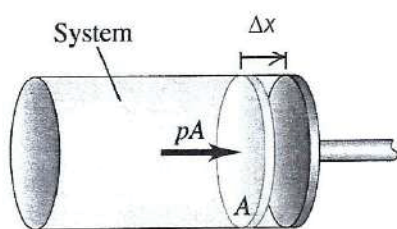


Fig. 11.7

Suppose that the constant pressure of the gas in the cylinder is p , the piston is of area A and the gas expands by pushing the piston a small distance Δx , then by the definition of work, the work done BY the gas is given by

$$W_{by} = F \Delta x = pA \Delta x$$

Since the change in volume is given by $\Delta V = A \Delta x$

$$W_{by} = p A \Delta x = p \Delta V$$

More generally, for varying pressure p , if the volume of gas changes from some initial V_i to some final V_f , the total work done BY the gas is then given by:

$$W_{by} = \int_{V_i}^{V_f} p dV \quad (\text{or area under the } pV \text{ graph})$$

11.7.2.2 $W_{on} = -W_{by}$

When a gas expands, the forces due to its pressure push back the piston. Hence work done BY the gas is positive. If energy is exchanged only between the gas and the surrounding and nothing else, we can interpret the positive work done BY a gas (by the gas pressure) when it expands, as negative work done ON the gas (by the surrounding pressure). It makes sense because the surrounding has withdrawn energy from the gas.

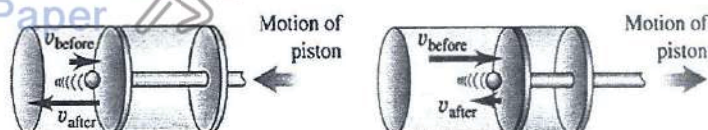
Likewise, when a gas contracts, the forces due to surrounding pressure push back the piston. Hence work done ON the gas is positive. We can interpret the positive work done ON a gas (by the surrounding pressure) when it contracts, as negative work done BY the gas (by the gas pressure). It makes sense because the gas has withdrawn energy from the surrounding.

We distinguish between the quantities work done ON the gas and work done BY the gas.

$$\text{work done ON the gas} = - (\text{work done BY the gas})$$

11.7.2.3 Microscopic View

When gas molecules hit an incoming piston, they rebound with speed larger than before the collision. The mean-square speed of the gas molecules, and thus the microscopic kinetic energy, increases. This agrees with the macroscopic interpretation that positive work is done on the gas when the gas is compressed.



Conversely, when gas molecules hit a retreating piston, they rebound with speed smaller than before the collision. The mean-square speed of the gas molecules, and thus the microscopic kinetic energy, decreases. This agrees with the macroscopic interpretation that negative work is done on the gas when the gas expands.

11.7.3 Increase in Internal Energy, ΔU

This term refers to the change in internal energy. For an ideal gas, the internal energy is proportional to its thermodynamics temperature ($U = \frac{3}{2}nRT$). Hence a positive ΔU must imply an increase in temperature of gas, and a negative ΔU must imply a decrease in temperature of gas.

11.7.4 $\Delta U = Q + W_{on}$

Putting all the terms together, you should realize that the **First Law of Thermodynamics is a statement of the Principle of Conservation of Energy** on a thermodynamic system. And there are two ways to raise the temperature of a system: (1) by heating and (2) by doing work on it, i.e. by compressing it.



gas.

Because solids and liquids have practically fixed volumes, the magnitude of work done is usually insignificant. This leads many of us to have this misconception that heating is the only way to raise the temperature of a system. But for gases, which are highly compressible, the amount of work is significant and plays an important role in changing the temperature of the



Likewise, many of us have the misconception that no change in temperature means no heat transfer. The truth is the temperature can remain constant if energy enters the gas by heat (Q is positive) and leaves through work (W_{on} is negative).

11.8 PV Graphs / Indicator Diagrams

The **state** of a fixed amount of gas is defined by three variables: Pressure P , Volume V , and Temperature, T . A thermodynamic process is said to have taken place when the gas changes from one state to another.

Changes to the pressure and volume of the gas are usually shown as lines on a graph of pressure against volume, called *PV* graphs. Each point on the graph corresponds to one state. The lines showing the transition in states during a thermodynamics process are called **indicator diagrams**.

For example, the *PV* graph in Fig. 11.8 shows different ways by which the state of an ideal gas may take in going from initial state 1 to final state 2.

1. The gas may undergo an isobaric process ($1 \rightarrow 3$) where the gas is held at constant pressure, followed by an isovolumetric process ($3 \rightarrow 2$) where the gas is held at constant volume.
2. Or the gas may undergo an isothermal process ($1 \rightarrow 2$) where the system is held at constant temperature. (It should be clear to you now that temperature of the gas is not directly shown on a *PV* graph. But it can always be calculated through the equation of state $PV = nRT$)

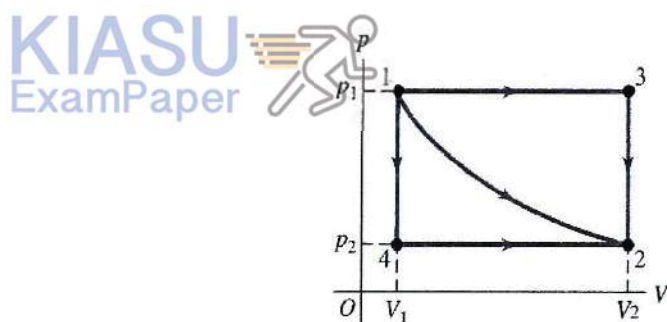


Fig. 11.8

11.8.1 Isothermal Process

An isothermal process is a process in which the temperature of the system remains constant. (i.e. $\Delta T = 0 \Rightarrow \Delta U = 0$ for a fixed mass of an ideal gas.)

An example of an isothermal process is the expansion of a cylinder of gas with very thin walls by pulling the piston extremely slowly, allowing sufficient time for the gas to gain heat through the thin wall to maintain thermal equilibrium with the surroundings.

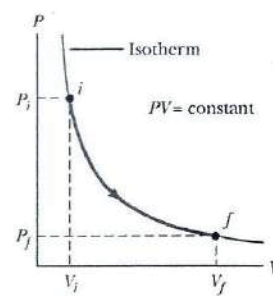
The indicator diagram of this isothermal expansion is a hyperbolic curve known as an **isotherm**.

By the First Law of Thermodynamics,

$$\Delta U = Q_{in} + W_{on}$$

$$0 = Q_{in} + \left(-\int_{V_i}^{V_f} p dV\right)$$

$$Q_{in} = \int_{V_i}^{V_f} p dV$$



For this example, the amount of heat gained from the surrounding ($Q > 0$) is equal to the area under the isothermal, which also represents the work done BY the gas ($W_{on} < 0$) expanding.

11.8.2 Isobaric Process

An isobaric process is a process that occurs at constant pressure. ($W_{on} = -P\Delta V$)

An example of an isobaric process is the expansion of a cylinder of gas with freely moving and frictionless piston, by heating the gas slowly, so that the pressure of the gas is always at equilibrium with the atmospheric pressure.

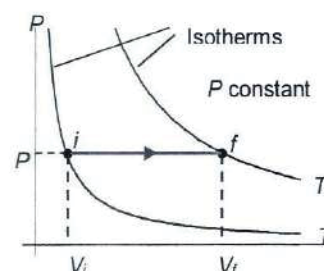
The indicator diagram for this isobaric expansion is a horizontal line.

By the First Law of Thermodynamics,

$$\Delta U = Q_{in} + W_{on}$$

$$= Q_{in} + (-p\Delta V)$$

$$= Q_{in} - p(V_f - V_i)$$



For this example, of the heat received by the gas ($Q > 0$); some of it is used to do work against the atmospheric pressure during the expansion ($W_{on} < 0$) and the rest goes to increasing the internal energy of the gas when it rises in temperature from T_i to T_f ($\Delta U > 0$).

11.8.3 Isovolumetric Process

An isovolumetric process is one that takes place at constant volume. ($W_{on} = 0$)

An example of an isovolumetric process is heating a cylinder of gas with the piston clamped at a fixed position.

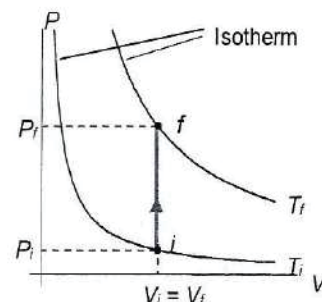
The indicator diagram for this isovolumetric expansion is a vertical line.

By the First Law of Thermodynamics,

$$\Delta U = Q_{in} + W_{on}$$

$$= Q_{in} + 0$$

$$\Delta U = Q_{in}$$



For this example, all the heat received by the gas ($Q > 0$) goes into increasing the internal energy of the gas when it rises in temperature from T_i to T_f ($\Delta U > 0$).

11.8.4 Adiabatic Process

An adiabatic process is one in which no energy leaves or enters the system by heat. ($Q = 0$)

An example of an adiabatic process is the sudden compression of a cylinder of gas by pushing the piston rapidly, so that there is negligible time for energy to transfer by heat out of the system. If the walls of the cylinder are well-insulated, it will further ensure minimal heat exchange during the compression.

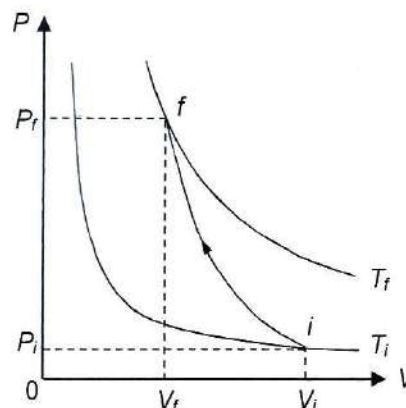
The indicator diagram for this adiabatic compression is a steep curve (steeper compared to the two isotherms it is bridging).

By the First Law of Thermodynamics,

$$\Delta U = Q_{in} + W_{on}$$

$$= 0 + \left(-\int_{V_i}^{V_f} p dV\right)$$

$$\Delta U = -\int_{V_i}^{V_f} p dV$$



For this example, all the work done on the gas ($W_{on} > 0$) (represented by the area under the graph) goes into increasing the internal energy of the gas when its temperature rises from T_i to T_f ($\Delta U > 0$).

11.8.5 Cyclic Process

A process that eventually returns a system to its initial state is called a cyclic process. As the internal energy U of the system depends only on the state of the system, in a cyclic process the net change of internal energy ΔU will be equal to zero.

By the First Law of Thermodynamics,

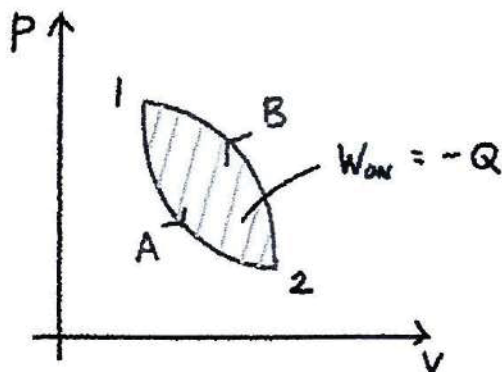
$$\Delta U = Q_{in} + W_{on}$$

$$0 = Q_{in} + W_{on}$$

$$W_{on} = -Q_{in}$$

So work done on the system is equal in magnitude but opposite in sign to the heat supplied to the system.

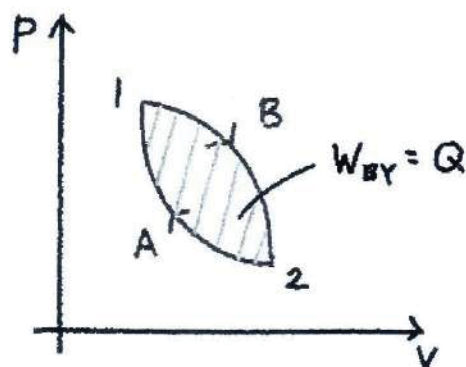
For example, the indicator diagram on the left shows the system returning to state 1 through processes A then B. The area under curve A represents work done by the system, whereas the area under curve B represents work done on the system. Since the latter is larger, the net work done ON the system during one complete cycle is positive ($W_{on} > 0$). For cyclic process, this also means that overall the system lost heat to the surrounding during one complete cycle ($Q < 0$).



In fact, since ΔU is zero, the area enclosed by the loop represents both the net work done ON the system, and the heat lost by the system, during one complete cycle.

The indicator diagram on the right shows the system returning to state 1 through processes B then A. Using similar logic, we can conclude that the net work done ON the system during one complete cycle is negative ($W_{on} < 0$). For a cyclic process, this also means that overall heat is supplied to the system during one complete cycle ($Q > 0$).

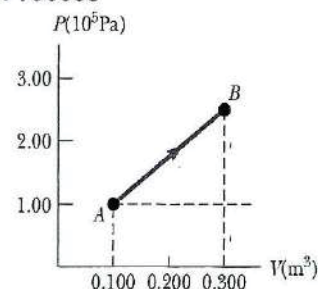
In fact, since ΔU is zero, the area enclosed by the loop represents both the net work done BY the system, and the heat supplied to the system, during one complete cycle.



Example 11.8.1 : Applying First Law of Thermodynamics to a General Process

A quantity of 4.00 moles of a monatomic ideal gas expands from an initial volume of 0.100 m^3 and pressure of $1.0 \times 10^5 \text{ Pa}$ to a final volume of 0.300 m^3 and pressure of $2.5 \times 10^5 \text{ Pa}$. Compute

- the work done on the gas,
- the change in internal energy of the gas, and
- the thermal energy transferred to the gas.



Solution:

- a) Since gas is expanding, $w_{on} < 0$.

$$\text{area under } P\text{-}V \text{ graph} = 3.50 \times 10^4 \text{ J}$$

$$w_{on} = -3.50 \times 10^4 \text{ J}$$

- b) Change in internal energy

$$\Delta U = U_{\text{final}} - U_{\text{initial}} = \frac{3}{2}(nRT_{\text{final}} - nRT_{\text{initial}})$$

$$= \frac{3}{2}(P_{\text{final}} V_{\text{final}} - P_{\text{initial}} V_{\text{initial}})$$

$$= \frac{3}{2}(2.5 \times 10^5 \times 0.300 - 1.0 \times 10^5 \times 0.100)$$

$$= 9.75 \times 10^4 \text{ J}$$

- c) Thermal energy transferred to the gas

$$\Delta U = w_{on} + Q_{in} \Rightarrow Q_{in} = \Delta U - w_{on}$$

$$Q = [9.75 - (-3.50)] \times 10^4$$

$$= 1.325 \times 10^5 \text{ J}$$

$$= 1.33 \times 10^5 \text{ J}$$

Example 11.8.2 : Applying First Law of Thermodynamics to a Cyclic Process

A gas is taken through the cyclic process described in the figure on the left. Find the net energy transferred to the system by heat during one complete cycle.

Solution:

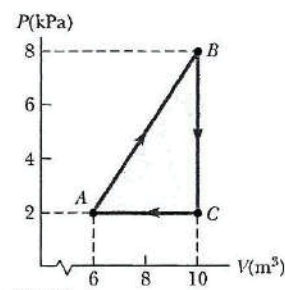
Area enclosed by the cyclic path ABC

$$= \frac{1}{2}(10-6)(8-2) = 12000 \text{ J}$$

Work done by gas AB > work done on gas (CA)

Thus net work done on gas = -12000 J

For cyclic process, $\Delta U = 0 \Rightarrow Q_{in} = -w_{on} = 12000 \text{ J}$



Appendix A: Necessary and Desirable Features in a Thermometric Property

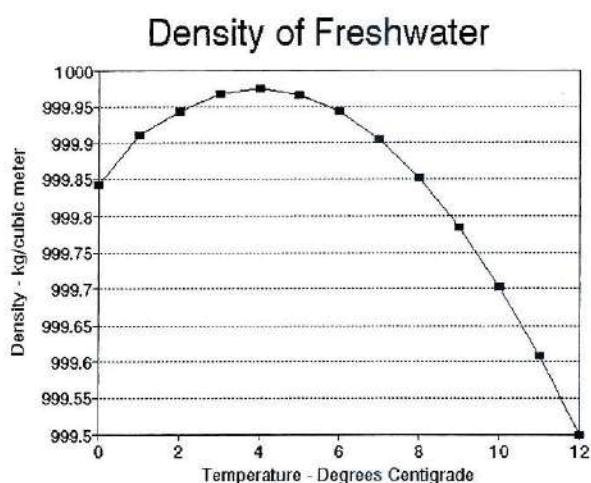
Necessary Features in a Thermometric Property

The thermometric property in a thermometer must satisfy the following characteristics:

- (1) *The property must vary continuously with temperature.*
- (2) *The property must be monotonic increasing or monotonic decreasing with temperature.*

This is so that the value of the thermometric property is unique for each temperature.

To illustrate what we mean, let us look at why water would not be a suitable liquid in a liquid-in-glass thermometer.



When water is cooled, it initially contracts and below 4 °C, it starts expanding as it is cooled further (Fig 8.2.1). This implies that for the same volume of water, there are two possible readings of temperature.

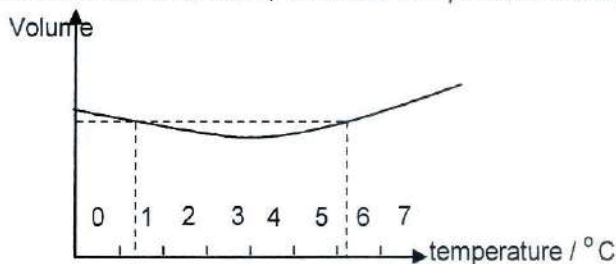


Fig 8.2.1 The variation of volume of water with temperature. Because of the unique behaviour of water with temperature, there will be a volume that corresponds to two temperatures that are different.

Pause and Think: However, if the range of temperatures to be measured is limited to above 4 °C, then the volume of water can be used as a thermometric property. Why is it so?

- (3) *The value of the property at a temperature must be reproducible.*

A thermometer is essentially useless, if on one occasion, the thermometric property gives one value and on another occasion, it gives another, for the same temperature. This implies that the same temperature measured on different occasions is indicated as different temperatures on the thermometer.

(4) The property used must be suitable over the range being measured.

For example, the working range of alcohol-in-glass thermometer is generally limited by the melting and boiling point of alcohol. Above the boiling point, alcohol would turn gaseous and below the melting point the alcohol would turn solid.

Other Desirable Features in a Thermometric Property and Thermometer: - ('Good To Have')

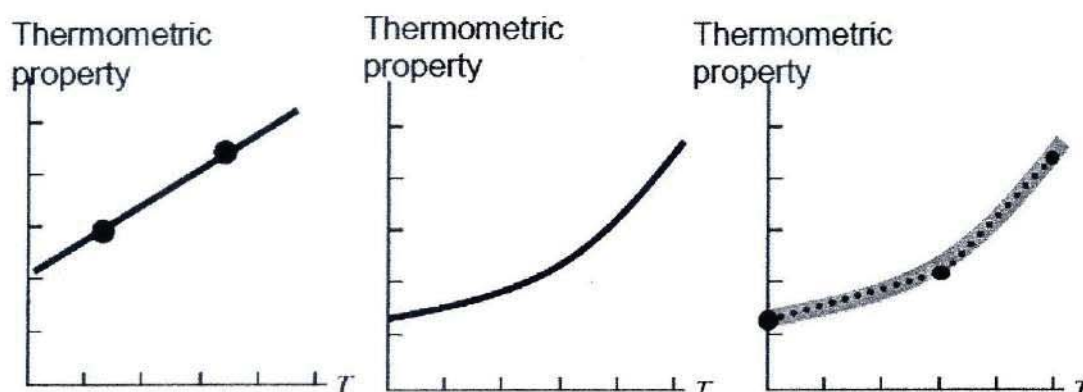
In addition to the essential features that a thermometric property must possess, there are also certain desirable features that we would like our thermometric property and thermometer to have. They are:

(1) Variation of the thermometric property is large or significant with temperature. (Good Sensitivity)

If variation of the property with temperature is not sufficiently large, we will not be able to detect changes of the property for small changes in temperature. This reduces the *sensitivity* (and *precision*) of the thermometer.

(2) Variation of thermometric property with temperature is linear.

A thermometric property that varies linearly with temperature T allows easy calibration for use in a thermometer. Using two known values of the property at just two temperatures allows the relationship between the property and temperature to be both interpolated and extrapolated.

**Fig. 1.3.2**

(a) Linear variation temperatures

(b) Non linear variation

(c) Calibration at 3 known for non-linear thermometric property

If the property is not linearly related to temperature, it can still be used. The thermometer will need to be calibrated at more fixed points. However, extrapolation of values beyond the range of the calibration points will be a challenge.

(3) The thermometric property should register changes in temperature quickly. (Good Responsiveness)

In this case, the *thermal conductivity* of the thermometer should be high, so that the thermometer can come to thermal equilibrium with the environment that it measures quickly.

(4) The thermometer should have a small heat capacity.

Having a *small heat capacity* implies that the thermometer will not need to absorb a significant amount of energy from the surroundings (or conversely not give off too much heat to the surroundings) for it to come to thermal equilibrium with the surroundings, so that it does not disturb the environment that it is measuring.

Appendix B: Types of Thermometers

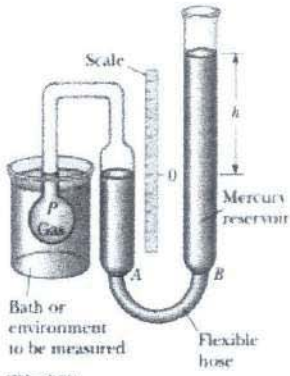
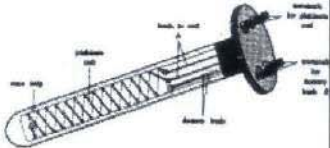
(You should have cover this in Year 4 / 'O' level Physics.)

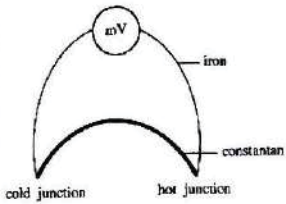
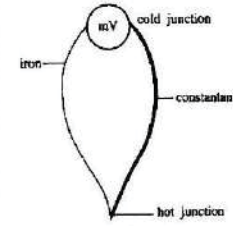
The choice of which thermometer to be used for a particular temperature depends on a variety of factors. These includes accuracy, sensitivity, convenience, size, nature of read-out required, cost, operating range of the thermometer, the temperature being measured, responsiveness and availability.

The different types of thermometers you should be familiar with are listed in the table below.

Type of Thermometer	Principle	Range/ $^{\circ}\text{C}$	Advantages (+) Disadvantages (-)
Liquid-in-glass thermometer 	Thermometric Property: Length of the liquid in the capillary. When the temperature rises, liquid expands and moves further up a capillary tube. Of all the liquid-in-glass thermometers, mercury is most commonly used. This is mainly because : (1) mercury expands almost uniformly with temperature; (2) mercury does not wet the glass; (3) mercury is a good conductor of heat and hence can come to thermal equilibrium very quickly with its surroundings; (4) mercury is opaque and can therefore be seen easily.	-39 to 500 (Depending on the liquid used in the glass. The operating range limited by the freezing and boiling point of the liquid used. Mercury freezes at about -39°C and boils at 357°C. Under normal pressure, the operating range of an ordinary mercury thermometer is between its freezing point to about 300°C. If the space above the mercury is filled with nitrogen or some other inactive gas, the exerted by the gas due will cause the boiling point of the mercury to rise, and such a thermometer can be used to above the point to approximately 500°C. For alcohol thermometers (with boiling point at 78°C) its operating range is approximately between -60°C and 130°C.)	+ Portable + Cheap, and simple to use + Direct reading - Not very accurate - Fragile - Responsiveness : Medium

[Table continues next page.]

Type of Thermometer	Principle	Range/ $^{\circ}\text{C}$	Advantages (+) Disadvantages (-)
Constant-volume gas Thermometer 	Thermometry Property : Pressure of the gas. The bulb contains a fixed mass of hydrogen, helium, or nitrogen gas. When the temperature rises, and a larger height difference h is needed in the manometer to keep the gas at fixed volume. (The open limb of the manometer can be raised to achieve this. The pressure of the gas is equal to ρgh plus atmospheric.)	-270 to 1500	+Very wide range +Very accurate and sensitive. - Very bulky and slow to respond, not suitable to varying - No direct reading - Mainly used as a standard to calibrate other more practical types types.
Platinum resistance Thermometer 	Thermometric Property: Resistance of the coil. When the temperature rises, the resistance of the platinum coil rises. The resistance is measured accurately with a circuit which makes use of balanced potential differences.	-200 to 1200	+ Wide range + Highly accurate, sensitive - Not direct reading - Slow in response; not suitable of varying temperatures.
Thermistor  Bead Thermistor encapsulated in glass	Thermometric Property: Resistance of the thermistor. When the temperature rises, the thermistor falls. The resistance is measured using circuit which can include a meter calibrated in degrees.	-25 to 150	+ Can be linked to other circuits or computer -Limited range - Not very accurate - Not lasting
Thermocouple:	Thermometric Property : Thermoelectric emf generated due to temperature differences. The circuit contains two junctions of dissimilar metals . When one junction is hotter than the other, a	-250 to 500	+Wide range +small, robust and compact +Fairly accurate +Quick Response + Widely used in industry for rapidly changing temperatures and

 Double Junction Thermocouple  Single Junction Thermocouple	small thermoelectric <i>emf</i> is generated. The magnitude of the <i>emf</i> depends on the temperature difference. It can include a meter calibrated in degrees.		temperatures at a 'point'. - Less accurate than constant volume gas and platinum resistance thermometer.
---	---	--	---

Type of Thermometer	Principle	Range/ $^{\circ}\text{C}$	Advantages (+) Disadvantages (-)
Pyrometer	Thermometric Property : Based on the different colours when object are heated to different high temperatures. Above temperatures of about 1700 K , all thermometers mentioned so far become quite useless, as the temperatures would have exceeded the melting points of the thermometers. Above this temperature, therefore, a method that is based on the property of the brilliance and colour that an hot object emits is used to measure its temperature instead. These thermometers are known as pyrometers and work on the basic principle of using the human eye to match the brightness of a hot object with a calibrated lamp filament inside the instrument. (Table below gives a guide to the temperature and comparative colour of a few commonly observed hot objects.)	above 1000	Only thermometer for above 1500°C. - Very judgmental and can only give an approximate range of the temperature an object is heated to.

Table : Guide to Colour and Approximate Temperature of Some Commonly Observed Hot Objects

Example	Colour	Approximate Temperature / K
Simmering hot-plate	Dull Red	900
Fully-on hot-plate	Red	1 000
Element of electric fire	Orange	1 100
Filament of electric lamp	Yellow	2 500
Surface of a red giant star	Yellow	3 300
Surface of the Sun	White	6 000
Surface of the white dwarf star (Sirius)	White	11 000
Surface of white giant star (Rigel)	Blue/White	25 000

Appendix C: Limitations of the Centigrade Scale

Key problems with the centigrade scale were :

(1) Two different types of thermometer using the two fixed points agree only at the calibration points.

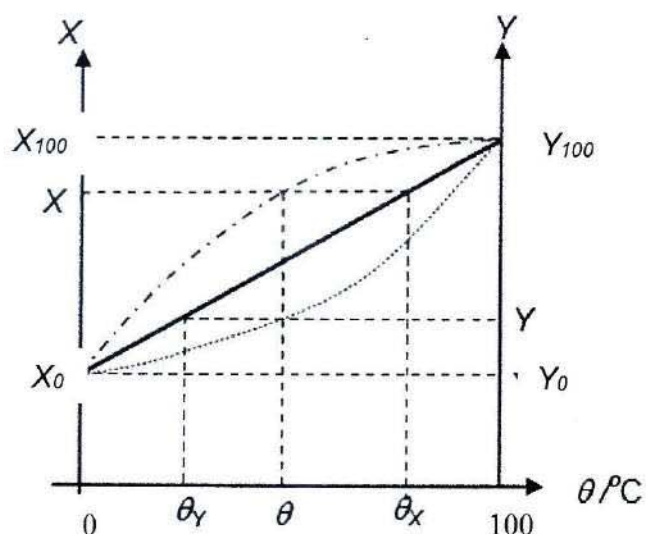
- Apart from the calibration points, a temperature measured with two different types of thermometers, e.g. a mercury-in-glass and a platinum resistance thermometer, may yield very different readings as different thermometric properties behave differently with temperature.

Let us use a graph (Fig. C1) to further illustrate what we mean.

Suppose we have two different types of thermometers, one having thermometric property X and the other having thermometric property Y , which we have calibrated on the centigrade scale.

Based on the assumption of linearity, we would expect both the variation of X and Y with temperature to be given by the middle straight line.

Note that the vertical scale for X and Y have been scaled so that both X_0 and Y_0 indicate 0°C . Similarly X_{100} and Y_{100} indicate 100°C . The thermometers agree at the calibration points.



———— Calibration curve for X and Y on the centigrade scale

- - - - - Actual Temperature response of X

..... Actual Temperature response of Y

Fig. C1. Disagreement between different thermometers.

Suppose we use these calibrated thermometers (X and Y) to measure the temperature of an object which has a "true" temperature of θ .

At θ , Thermometric property registered by thermometer $X = X_\theta$

$\Rightarrow$ Temperature "read" by thermometer $X = \theta_X$

Thermometric property registered by thermometer $Y = Y_\theta$

$\Rightarrow$ Temperature "read" by thermometer $Y = \theta_Y$

Note that

- Both thermometers do not indicate the true temperature θ , as both properties do not behave linearly with temperature.
- They also do not agree with each other ($\theta_X \neq \theta_Y$) as they do not respond similarly with temperature.

(2) Disagreement between Instruments of the Same Type

- Even two thermometers of the same type may not agree with each other as the behaviour of the thermometric property with temperature depends on many external factors.
- For example, it is not possible to manufacture two thermometers of the same type to be exactly the same. In a mercury-in-glass thermometer it is extremely difficult to manufacture a bore of uniform cross-sectional area and hence the linear expansion between thermometers may differ slightly.

(3) Problems in Reproducibility of the Fixed Points

- The ice point and steam point of water are actually very difficult to duplicate in the laboratory.
- For ice point,
 - one cannot simply put in some ice cubes in water. An ice cube in water may have a temperature of 0°C only right at the surface of the ice cubes, while the surrounding water a short distance away from the ice cubes may be at a temperature of 2°C .
 - To obtain 0°C (melting point of ice), the ice water mixture must consist of finely crushed ice that is kept in equilibrium with the water by continuously stirring.
 - Melting point of ice instead of the freezing point of water is used as the fixed point. This is because water does not necessarily freeze at 0°C . Pure water under very clean conditions may go well below 0°C before suddenly crystallizing. Ordinary tap water is usually sufficiently contaminated so that it freezes below 0°C .
- For boiling point of water, it depends very much on the atmospheric pressure. For example, boiling point is defined to be 100°C at normal atmospheric pressure at sea level. At normal atmospheric pressure in Denver, about a mile above sea level, the boiling point is about 95°C .

Appendix D: Calibration of a Thermometer Using the Ideal Gas Scale

The following steps illustrate how temperature is quantified in an ideal-gas scale :

Step 1 : Choose a Thermometric Property, X

We have seen that when the temperature of an ideal gas changes, both the pressure of the ideal gas as well as the volume of the ideal gas can change. Therefore, the property of the ideal gas that is used to set up the new scale of temperature is the product of pressure p and volume V i.e. pV .

Charles and Guy-Lussac found that a linear relationship exist between pV and the temperature. Hence we have

$$pV = aT + b$$

where a and b are constants in the experiments.

The unit of temperature measured on this scale is Kelvin (K).

Step 2 : Choosing Two Fixed Points

Lower Fixed Point : Absolute Zero

As we have seen from Charles and Guy-Lussac's experiments that the graph of the product of pressure and volume curve is extrapolated to very low temperatures, all the gases seems to behave similarly and intersect the temperature axis at -273.15°C . Hence, if we choose this point to be zero on our new temperature scale, the pV axis will be shifted to the left as in Figure B1.

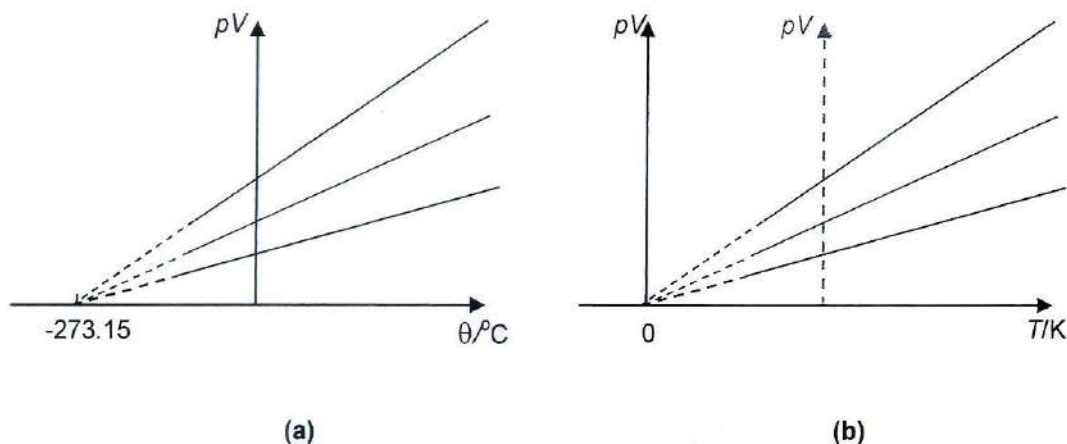


Figure B1. (a) The pV of gases at extreme low pressures and volumes all extrapolates to the value of -273.15°C . (b) By assigning -273.15°C to be 0 K on the new temperature scale, the pV axis is shifted to the right, and pV becomes proportional to T , the temperature on the new temperature scale.



With the lower fixed point chosen to be $T = 0 \text{ K}$ when $pV = 0$,⁴⁴

$$pV = aT + b \quad \Rightarrow \quad b = 0$$

Hence, $pV = aT$

Therefore, we see that the thermometric property pV varies proportionally with temperature T .

To determine, the constant a , we need to define a second fixed point, that is temperature at triple point of water, 273.16K.

Upper Fixed Point : Triple Point of Water

The second fixed point used is the **triple point of water**. The triple point of water was chosen to be the upper fixed point on the ideal gas scale as it could be precisely reproduced. It was found to occur at 0.01°C and hence was designated a value of 273.16 K on the ideal gas scale so that there would still be 100 divisions between steam and ice point.

We use a triple point of water cell, as shown in Fig. 1.5, to calibrate thermometers. It consists of a cylinder of borosilicate glass with a reentrant tube serving as a thermometer well, filled with high-purity, gas-free water, and sealed. The cell is cooled until ice, water, and water vapour are all present and in equilibrium. The temperature is then at triple point, 273.16 K by definition. A thermometer may be calibrated by inserting it into the central tube.

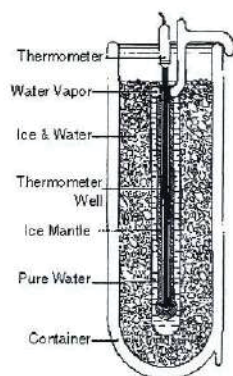


Fig. 1.5a Triple point of water cell

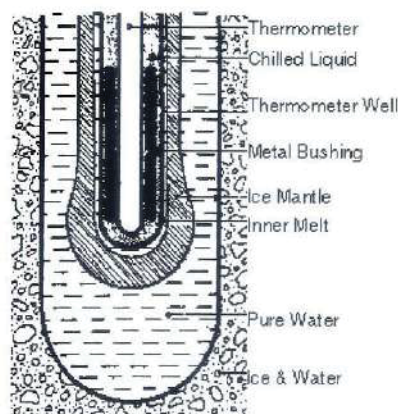


Fig. 1.5b Cross section of triple point of water cell

⁴⁴ Note that the point 0 K is a point that is obtained from the extrapolation of the pV graph. In reality, the temperature of absolute zero (0 K) cannot be achieved, although laboratory experiments incorporating the laser cooling of atoms have come very close. What would happen to a gas if its temperature could reach 0 K (and did not liquefy or solidify)? As the pV vs. T curve indicates, the pressure exerted on the wall becomes zero. In a later section, we shall show that this pressure of the gas is proportional to the average kinetic energy of the gas molecules. Thus according to classical physics, the kinetic energy of the gas molecules would become zero at absolute zero, and molecular motion would cease; hence, the molecules would settle out on the bottom of the container. Quantum theory modifies this prediction and shows that some residual energy, called the *zero-point energy*, would remain at this low temperature.

Appendix E: Getting the Ideal Gas Law

The early experiments in the 17th century and 18th century with ideal gases resulted in three laws known as the **gas laws**. They are, in chronological order :

- o Boyle's Law
- o Charles' Law
- o Pressure Law

Boyle's Law states that for a fixed mass of gas at constant temperature, the product of the pressure P of the gas and the volume V occupied by the gas is a constant.

From Boyle's Law, we have

$$PV = \text{constant} \Rightarrow P = \frac{\text{constant}}{V}$$

Hence, if we were to plot a graph of pressure P versus volume V , we will obtain Fig. E1.

Charles' Law states that for a fixed mass of gas at constant pressure P , the volume V occupied by the gas is directly proportional to the absolute temperature of the gas.

From Charles' Law we have,

$$V \propto T \Rightarrow V = (\text{constant})T$$

Hence, a plot of volume V versus temperature T would give Fig. E2.

Lastly, we have the Pressure Law which was also known as the Guy-Lussac's Law.

The **Pressure Law** states that for a fixed mass at constant volume, the pressure P of the gas is proportional to the absolute temperature T the gas is at.

Therefore, we have

$$P \propto T \Rightarrow P = (\text{constant})T$$

and a plot of pressure P against temperature T will give Fig. E3.

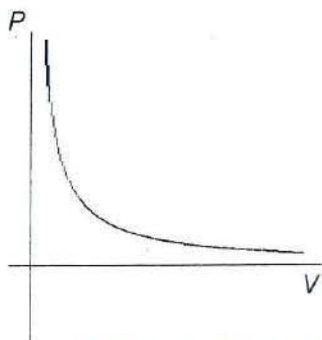


Fig. E1. Boyle's Law. Graph of pressure P vs. volume V of a fixed mass of ideal gas held at constant temperature T .

$$P = (\text{constant}) \frac{1}{V}$$

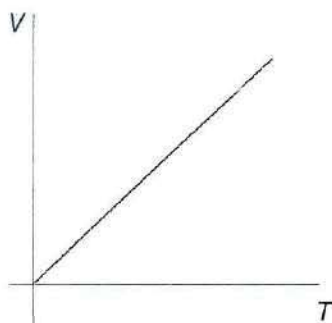


Fig. E2. Charles' Law. Graph of volume V vs. temperature T of a fixed mass of ideal gas held at constant pressure P .

$$V = (\text{constant})T$$

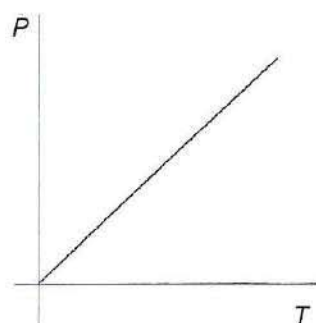


Fig. E3. Pressure Law. Graph of pressure P vs. temperature T of a fixed mass of ideal gas held at constant volume V .

$$P = (\text{constant})T$$

To derive the equation of state, we assume that we have a fixed mass of ideal gas in a cylinder fitted with a smooth light piston as shown in Fig. E4.

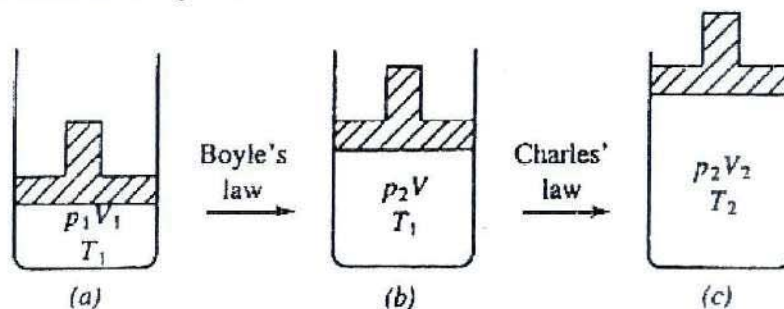


Fig. E4

If the pressure of the gas is changed to p_2 at constant temperature, according to Boyle's law, the volume V of the gas is governed by the equation

$$p_2 V = p_1 V_1 \dots\dots\dots(1)$$

If the temperature of the gas is then changed to T_2 at constant pressure, according to Charles' law, the new volume V_2 is given by

$$V_2 / T_2 = V / T_1 \dots\dots\dots(2)$$

(1) X (2):

$$\begin{aligned} (p_2 V_2) / T_2 &= (p_1 V_1) / T_1 \\ \Rightarrow pV / T &= \text{constant} \end{aligned}$$

The value of the constant depends on the number of moles of gas n as well as the molar gas constant R and we have:

$$pV = nRT$$

which can be verified experimentally.

This equation is the **equation of state for the Ideal Gas** and is also known as the **Ideal Gas Law**.

H2 Physics (Syllabus 9749)

Content

- Concept of an electric field
- Electric force between point charges
- Electric field of a point charge
- Uniform electric fields
- Electric potential

Learning Outcomes

Candidates should be able to:

- show an understanding of the concept of an electric field as an example of a field of force and define electric field strength at a point as the electric force exerted per unit positive charge placed at that point.
- represent an electric field by means of field lines.
- recognise the analogy between certain qualitative and quantitative aspects of electric field and gravitational field.
- recall and use Coulomb's law in the form $F = \frac{Q_1 Q_2}{4\pi\epsilon_0 r^2}$ for the force between two point charges in free space or air.
- recall and use $E = \frac{Q}{4\pi\epsilon_0 r^2}$ for the field strength of a point charge in free space or air.
- calculate the electric field strength of the uniform field between charged parallel plates in terms of the potential difference and plate separation.
- calculate the forces on charges in uniform electric fields.
- describe the effect of a uniform electric field on the motion of charged particles.
- define the electric potential at a point as the work done per unit positive charge in bringing a small test charge from infinity to that point.
- state that the field strength of the electric field at a point is numerically equal to the potential gradient at that point.
- use the equation $V = \frac{Q}{4\pi\epsilon_0 r}$ for the electric potential in the field of a point charge, in free space or air.

Contents

Pg

H2 Physics Syllabus 9749	2
12.1 Review of Basic Concepts	4
12.2 The Electric Field	6
12.2.1 Electric Field Strength.....	6
12.2.2 Electric Field Lines	8
12.2.3 Force between Two Point Charges (Coulomb's Law)	10
12.2.4 Electric Field around a Point Charge	12
12.3 Electric Potential	14
12.3.1 Potential due to a Point Charge.....	14
12.3.2 Electric Potential Energy.....	15
12.4 Relation between Potential Energy and Electrical Force	19
12.4.1 Relation between Potential Gradient and Electric Field Strength.....	19
12.5 Representation of Electric Fields by Equipotential Lines or Surfaces	20
12.6 Electric Field between Two Charged Parallel Plates	23
12.6.1 Motion of Charged Particles in a Uniform Electric Field.....	23
12.7 Analogy between Electric fields and Gravitational Fields	26
Appendices	27
Appendix A: Deriving the Potential of an Electric Field by a Point Charge.....	27
Appendix B: Deriving the Relationship between Force and Potential Energy.....	28

References

- Jewett, J. W., & Serway, R. A. (2008). *Physics for Scientists and Engineers*. (7th ed.). Belmont, USA: Thomson Learning, Inc.
- Young, H. D., & Freedman, R. A. (Eds.). (2007). *University Physics with Modern Physics* (12th Edition ed.): Addison Wesley.
- Duncan, T. (2000). *Advanced Physics*. (5th ed.). London: John Murray (Publishers) Ltd.

Attempts have been made by the tutors to keep the information presented here as accurate as possible. Nevertheless, there could still be errata present and we seek your understanding in this matter. We welcome all feedback and comments to help us further improve the notes. For comments and feedback on this set of lecture notes, please kindly contact your Physics Tutor.

12.1 Review of Basic Concepts

Electrical Charges

Many of us could have encountered such an experiment in our secondary school years. When a glass rod that has been rubbed with a silk cloth is brought near another rubber rod which has been rubbed with fur, they attract each other. However, when two rubber rods which have been rubbed with fur are brought together, they repel each other (Fig. 1).

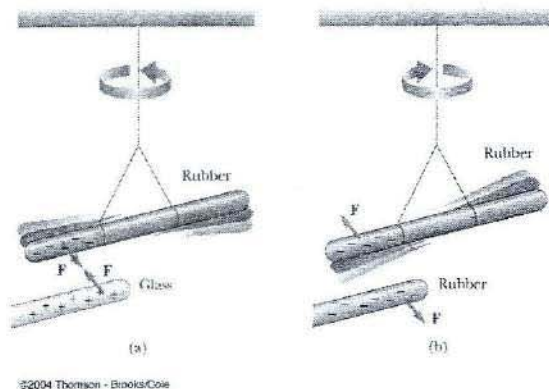


Fig. 1. (a) A negatively charged rubber rod is attracted to a positively charged glass rod. (b) A negatively charged rubber rod is repelled by another negatively charged rubber rod. {Jewett, 2008}

Using similar experiments like the one described above, Benjamin Franklin (1706 – 1790) showed that:

- There are two kinds of charges - **positive charge and negative charge**. (It was later discovered that electrons carry negative charge while protons carry positive charge)
- Charges of the same sign repel one another. Charges with opposite signs attract one another.

Atomic Structure

All materials are made up of atoms. The structure of atoms can be described in terms of three particles – the negatively charged electron, the positively charged proton and the neutral neutron.

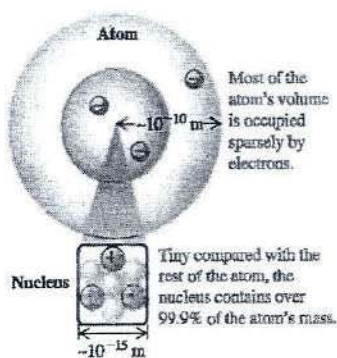


Fig. 2. The structure of an atom. The particular atom depicted here is lithium. (Young & Freedman, 2007)

The protons and neutrons in an atom make up a small, very dense core called the **nucleus**, which is approximately 10^{-15} m in size (Fig. 2). Surrounding the nucleus are electrons, extending out to distances of approximately 10^{-10} m from the nucleus. These negatively charged electrons orbit the nucleus and the attractive electrical force on the electrons by the protons provides the force required to keep the electrons in orbit.

The **electrical charge** and the **mass** of the electrons, protons and neutrons are given in the table below:

	Charge	Mass
Electron	$-e = -1.60 \times 10^{-19} \text{ C}$	$9.11 \times 10^{-31} \text{ kg}$
Proton	$+e = 1.60 \times 10^{-19} \text{ C}$	$1.67 \times 10^{-27} \text{ kg}$
Neutron	No charge (0 C)	$1.68 \times 10^{-27} \text{ kg}$

*e is known as the fundamental charge or the elementary charge.

Neutral Atoms, Positive Ions and Negative Ions

A **neutral atom** is one in which the **number of electrons equals the number of protons**. If one or more electrons are removed from this atom, the atom becomes positively charged and is known as a **positive ion**. A **negative ion** is an atom that has *gained* one or more electrons. This process of gaining or losing electrons is called **ionisation**. (Fig. 3 illustrates an example.)

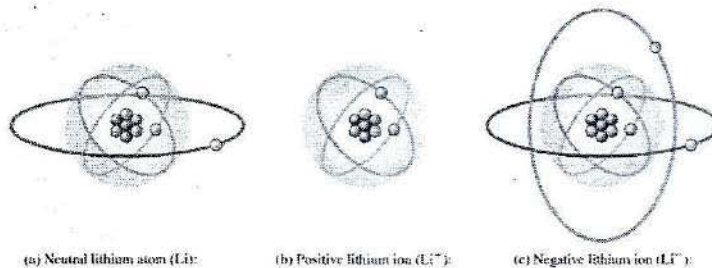


Fig. 3. (a) A neutral lithium atom has equal number of protons and electrons. (b) A positive lithium ion has a deficit of electrons. (c) A negative ion has an excess of electrons. (Young & Freedman, 2007)

Furthermore, when the total number of protons in body equals the number of electrons in the body, the total charge is zero and the body is on the whole electrically neutral. **To charge a body *negatively*, we add electrons** to the originally neutral body so that it has excess negative charges. Similarly, **to positively charge a body**, we can **remove electrons** (or sometimes add protons). Hence, when we speak of *charge of a body*, we always mean its *net charge*.

Principles of Conservation and Quantization of Charges

Implicit to the discussion above are two very important principles:

- Principle of Conservation of Charges
- Principle of Quantization of Charges

Principle of Conservation of Charges:

The *principle of conservation of charges* states that charges cannot be created or destroyed. Hence, for any closed system, the sum of all electric charges must be constant.

Suppose we rub a plastic rod with a piece of fur, both initially uncharged, and the plastic rod acquires a positive charge and the fur a negative charge. The principle of conservation of charges tells us that the positive charge acquired by the rod must be equal in magnitude to the charge on the fur. The rod has acquired a positive charge as it has transferred its electrons to the fur. Hence, in the charging process, charge is not created or destroyed; it is merely *transferred* from one body to another.

Principle of Quantization of Charges

In 1909, Robert Millikan (1868 – 1953) discovered that **electric charge always occurs as integral multiples of a fundamental amount of charge e** which is also known as the **elementary charge**. Charges are therefore said to be **quantized**, i.e. any charge q must be equal to $\pm Ne$ where N is some integer. (See http://en.wikipedia.org/wiki/Oil_drop_experiment.)

12.2 The Electric Field

How do objects not in contact with each other affect each other's behaviour? How does a proton exert a pull on an electron at a distance? To explain 'action at a distance', physicists developed the idea of a field. You have learnt about the gravitational field that results from the presence of a mass. In this chapter you will learn about the electric field, which comes about due to the presence of a net charge.

The electric force between charged objects is an example of action at a distance. One charged object exerts a force on another even though they are not in physical contact. A charged object **A** sets up an electric field in the space around it. When a second charged object **B** is in the field set up by **A**, object **B** experiences a force (Fig. 4). Just as object **B** experiences a force as it is in the field set up by **A**, it also generates an electric field around itself and hence **A** experiences a force by **B** as it is in the electric field set up by **B**.

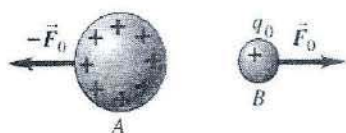


Fig. 4. When the charged object B is placed in the electric field set up by charged object A, it experiences a force. Similarly, A experiences a force as it is in the field of B. (Young & Freedman, 2007)

Electric Field

An electric field is a region of space such that when a charge is placed at a point in the region, it will experience an electrical force.

12.2.1 Electric Field Strength

Definition: Electric Field Strength, E

The electric field strength¹ at a given point is the **force per unit charge** that acts on a small **positive test charge** placed at that point.

$$E = \frac{F}{q}$$

- Electric field strength is a **vector quantity**.
- The *direction of the electric field* at a point is the **direction of the electric force** experienced by a **positive test charge** if placed at that point.

The S.I. unit of electric field strength is newton per coulomb (N C^{-1}).

¹ Note that though we distinguish the terms *electric field strength* and *electric field* here, they are often used interchangeably.

Force on a charge q due to an Electric Field E

From the definition of electric field strength E , if a charge q is placed in an electric field where the field strength is E , the field will exert a force F on the charge given by

$$F = qE$$

Notes:

From the equation above we see that:

- the **magnitude of the force** on the charge placed in the field is given by $F = |qE|$
- The **direction of the force on the charge is the same as that of** the electric field at the point if the charge **q is positive**.
- The **direction of the force on the charge is opposite to that of** the electric field if the charge **q is negative**.

Example 1

An isolated electron (in a vacuum) near the surface of the earth was found to be accelerating directly away from the surface with an acceleration $a = 2.63 \times 10^6 \text{ m s}^{-2}$.

[elementary charge, $e = 1.60 \times 10^{-19} \text{ C}$, mass of the electron, $m_e = 9.11 \times 10^{-31} \text{ kg}$]

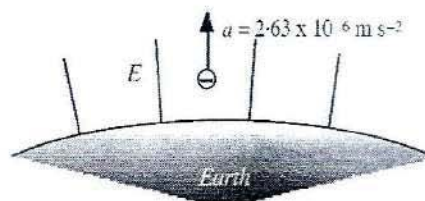


Fig. 5

- What is the direction of the electric field near the surface of the earth at its location?
- Calculate the magnitude of the electric field near the surface.

a) The force is acting upwards. The charge is $-e$. The direction of the field is downwards.

b) From the definition of field strength, $F_E = qE$

From N2L, $F_{\text{net}} = ma$

Ignoring gravity $\Rightarrow F_{\text{net}} = F_E$

This gives $ma = qE$, or $E = \frac{mq}{q}$

Magnitude of field $E = \frac{mq}{e} = 1.5 \times 10^5 \text{ N C}^{-1}$

Note that acceleration $\gg g$!

12.2.2 Electric Field Lines

A convenient way of visualizing electric field patterns is to draw straight or curved lines that are parallel to the electric field vector at every point in space. These **field lines** are also known as **lines of force** as they represent the direction of the force experienced by a *positive charge* placed at that point in the field.

In general, the electric field lines must be able to represent the variations in the **magnitude** and **direction** of the electric field in a region of space, which is done in the following manner:

○ **Magnitude of the Strength of the Electric Field:**

- The *number of field lines per unit area*, through a surface perpendicular to the lines, is proportional to the *magnitude* of the electric field.
- Hence, where the lines are close together, the electric field is strong, while where the lines are far apart, the field is weak (Fig. 6).

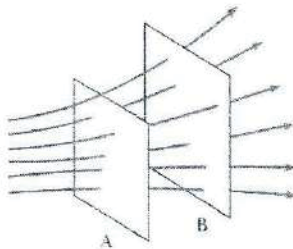


Fig. 6. Electric field lines penetrating two surfaces. The magnitude of the field is greater on surface A than on surface B. (Jewett & Serway, 2008)

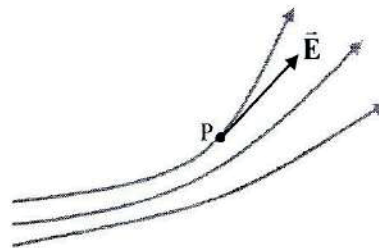


Fig. 7. The electric field at point P is tangent to the field line through P.

○ **Direction of the Electric Field:**

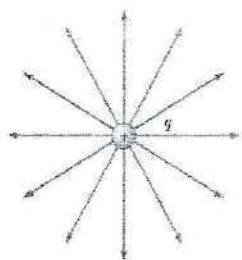
- The **direction of the electric field strength E at a point in an electric field is tangent to the electric field line at that point** and is the same as that of the electric field (Fig. 7).

Rules for Drawing Electric Field Lines²:

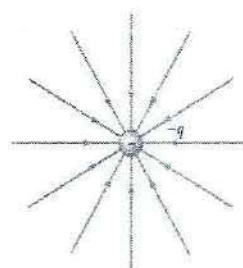
- i. Electric field lines must begin from a positive charge and end on a negative charge. In the case of an excess of one type of charge, some lines will begin or end infinitely far away.
- ii. The number of lines drawn leaving a positive charge or approaching a negative charge is proportional to the magnitude of the charge.
- iii. Field lines do not intersect because E at a point can only point in one direction.

² There are also rules associated with drawing field lines from equipotential lines. We will cover that in a later section and summarize all the rules there.

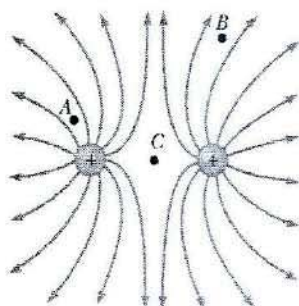
Fig. 8 (a) to (f) illustrates the field patterns of some common charge systems.



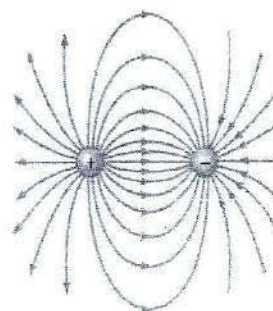
(a) Single positive charge



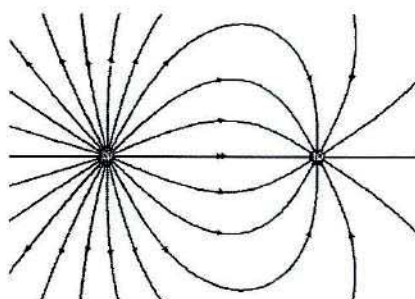
(b) Single negative charge



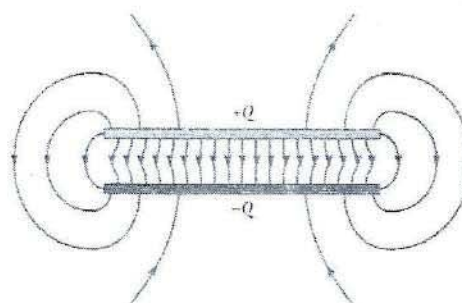
(c) Two positive charges of equal magnitude



(d) Two opposite charges equal in magnitude



(e) Two opposite charges, the magnitude of one charge being twice the magnitude of the other



(f) A pair of oppositely-charged parallel plates³

Fig. 8. Electric field of some common charge systems. (Jewett & Serway, 2008)



Electric field lines are an effective way to pictorially represent the direction and strength of an electric field at different points in a region of space around a charged object. Only a finite number of lines from each charge can be drawn, which makes it appear as if the field only exists in certain parts. The field, in fact, is continuous, existing at every point in the region around the charged object. Also try to avoid obtaining the wrong impression from a two-dimensional drawing of field lines used to describe a three-dimensional situation.

³ The charged parallel plates are said to produce a uniform electric field between them (indicated by the parallel lines between the plates), except near the edges, where there are *fringe effects*. This is because the charges at the edges are not distributed uniformly, unlike the other charges that are nearer to the centre of the plates.

Charge system (d) is known as an electric dipole and is a commonly occurring configuration, for instance, the charge distribution between the oxygen atom and hydrogen atom in a water molecule constitutes an electric dipole.

12.2.3 Force between Two Point Charges (Coulomb's Law)

As stated earlier, when a charge Q is placed near another charge Q' , each of them will experience a force as they are in the field set up by the other charge. If Q and Q' are point charges, the magnitude of the force can be determined by **Coulomb's Law**.

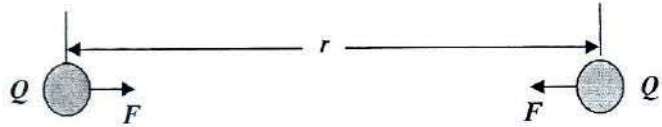


Fig. 9. The force F between two point charges is determined by Coulomb's Law.

Coulomb's Law

The magnitude of the electrical force acting between two point charges is proportional to the product of the magnitude of the charges and inversely proportional to the square of the distance between them.

$$F = \frac{1}{4\pi\epsilon_0} \frac{QQ'}{r^2}$$

where F : magnitude of the force between charges Q and Q'
 ϵ_0 : permittivity of free space, and
 r : the distance between Q and Q'

Notes:

- There are, in reality, one force F acting on Q and another force F' acting on Q' . These two forces are actually a (*Newton's Third Law*) *action-and-reaction pair*.
- Permittivity:**

In the S.I. system of units, the constant of proportionality in Coulomb's Law is usually written as $1/4\pi\epsilon$. The constant ϵ is known as *permittivity*. Permittivity is a property of the medium, i.e. its value depends on the medium in which the charges are placed. If the space between the charges is '*free space*' (or *vacuum*), its permittivity (known as permittivity of free space) will be denoted as ϵ_0 .

$$\epsilon_0 = 8.85 \times 10^{-12} \text{ F m}^{-1}$$

where F (farad), is the unit for capacitance (expressed in SI units). $1\text{F} = 1\text{CV}^{-1}$.



For most practical purposes the permittivity of 'air' is the same as that of 'free space'.

When using Coulomb's Law, just **substitute the magnitude of the charges** (without the signs) to find the **magnitude of the force**.

Indicate the direction of the force (or whether it is attractive or repulsive) separately.

KIASU
ExamPaper

The Principle of Superposition (for electrical forces)

What happens if there are more than two charges?

The force between any two charges in a group of charges is independent of the presence of the other charges. Hence, when more than two charges are present, the force between any pair of them is still governed by Coulomb's law.

The resultant force on any one of them is equal to the vector sum of the forces exerted by the other individual charges.

For example, if four charges are present, the resultant force exerted by particles 2, 3 and 4 on particle 1 is given by

$$\vec{F} = \vec{F}_{12} + \vec{F}_{13} + \vec{F}_{14}.$$

Example 2:

In a water molecule, the oxygen atom acts as a negative point charge of magnitude 1.1×10^{-19} C and the hydrogen atoms act as positive point charges of magnitude 0.55×10^{-19} C. The distance r between each hydrogen atom and the oxygen atom is 1.0×10^{-10} m.

Calculate the net force due to the two hydrogen atoms experienced by the oxygen atom.

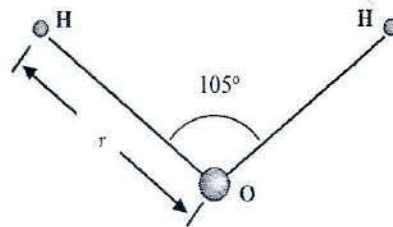


Fig. 10. Figure for Example 2.

$$F_{\text{net}} = F_{12} + F_{13}$$

Because of symmetry, horizontal forces cancel

Because of symmetry,

$$F_{\text{net}} = F_{12,y} + F_{13,y}$$

$$= 2 F_{12,y}$$

$$= 2 F_{12} \cos \alpha$$

$$= 2 \left(\frac{Q_1 Q_2}{4\pi\epsilon_0 r^2} \right) \cos 52.5^\circ$$

$$= 6.62 \times 10^{-19} \text{ N, vertically upwards}$$

12.2.4 Electric Field around a Point Charge

With the electric field strength E defined, and *Coulomb's Law*, we can now quantitatively describe the electric field around a point charge.

Consider an isolated point charge Q that is placed in free space. An electric field is set up by Q , in the space around it. If a small positive charge q is brought into the space and is placed at a distance r from Q , a force F will be exerted by the field of Q on the charge q .

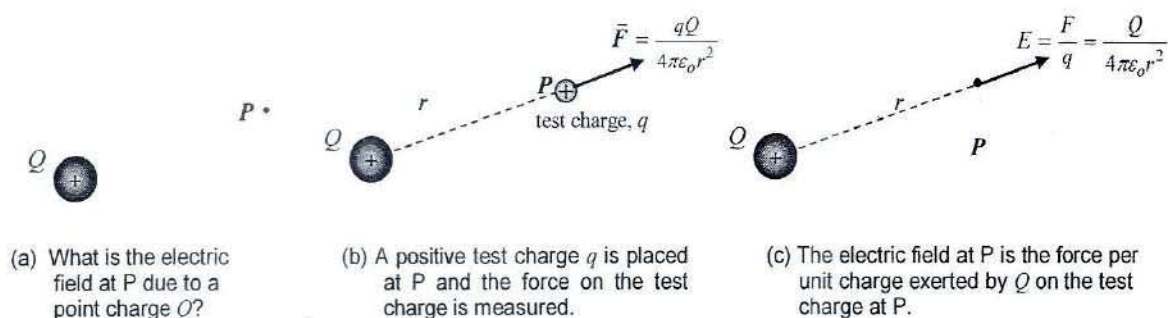


Fig. 11. Determining the electric field of a point charge

According to Coulomb's law, the magnitude of repulsion or attraction between Q and q is

$$F = \frac{1}{4\pi\epsilon_0} \frac{Qq}{r^2}$$

Hence, the **magnitude of the electric field strength of a point charge Q** at a distance r away from the point charge,

$$E = \frac{F}{q}$$

$$\Rightarrow E = \frac{1}{4\pi\epsilon_0} \frac{Q}{r^2}$$

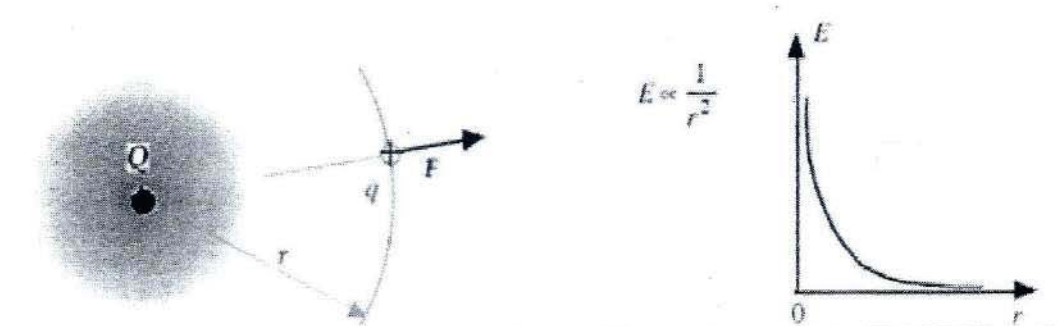


Fig. 12. Electric field of a point charge Q

Notes:

- E is a vector quantity, and its direction at a point is given by the direction of the force experienced by a positive charge if it is placed at that point.
- The field is radial for a point charge. It is directed uniformly in all directions outward from the centre if Q is a positive charge and inward toward the centre if Q is a negative charge. At all points that are an equal distance away from Q the magnitude of E is the same.



One can just substitute the absolute value of Q in the equation to calculate the magnitude of the field strength and consider the direction separately.

Principle of Superposition (for Electric Fields)

The Principle of Superposition can be used to compute the resultant electric field strength at a point in the combined field produced by a system of charges.

The resultant electric field $\vec{E}$ at a point P in an electric field is the vector sum of the fields at P due to each point charge in the system.

$$\vec{E} = \vec{E}_1 + \vec{E}_2 + \vec{E}_3 + \dots$$

Example 3

A charge $+1.4 \mu\text{C}$ and a charge $+2.6 \mu\text{C}$ are separated at 15 cm from each other. At what point P , other than a point at infinity, is the electric field zero? This point is known as the neutral point in the field between the two charges.

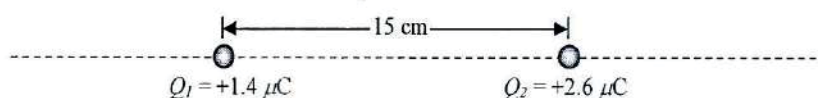


Fig. 13. Figure for Example 3

Let P be a distance x to the right of charge Q_1 .

At the neutral point, the magnitude of the field due to Q_1 is equal to the magnitude of the field due to Q_2 .

$$\frac{Q_1}{4\pi\epsilon_0 x^2} = \frac{Q_2}{4\pi\epsilon_0 (0.15 - x)^2}$$

$$(0.15 - x)\sqrt{Q_1} = x\sqrt{Q_2}$$

$$x = \frac{0.15\sqrt{1.4 \times 10^{-6}}}{\sqrt{1.4 \times 10^{-6}} + \sqrt{2.6 \times 10^{-6}}} = 0.0635 \text{ m} = 6.35 \text{ cm}$$

12.3 Electric Potential

Using the concept of energy, it is possible to associate a scalar quantity with every point in an electric field.

Suppose an external agent does work, W , in bringing a small positive charge q from infinity to a particular point in an electric field.

Definition: Electric Potential, V

The electric potential V at a point in an electric field is defined as the work done per unit charge, by an external force, in moving a small positive test charge from infinity to that point in the electric field, without any change in the kinetic energy.

$$V = \frac{W}{q}$$

The SI unit of potential is J C^{-1} but it is more common to use the volt, V.

Notes:

- Electric potential is a scalar quantity.
- Note that the test charge moved is not the one generating the field, it is the charge used to test the field.



The work done in this definition refers to the work done by the external force and is commonly referred to as the "work done on the system". Some questions in A-levels will ask you to look for "work done by the system" instead – it is simply the work done by the internal force that the external force is trying to overcome. You can simply find it by using:

$$\text{Work done on system (by external force)} = - \text{Work done by system (by the associated field force)}$$

12.3.1 Potential due to a Point Charge

The electric potential in the field by a point charge⁴ Q is given by

$$V = \frac{1}{4\pi\epsilon_0} \frac{Q}{r}$$

where V is the potential at a point the electric field, at a distance r away from the point charge.

Note:

What if I have two or more point charges, how do I determine the potential at a point in the electric field set up by these point charges?

Since the electric potential is a scalar quantity, the resultant potential at a point in the field is the **algebraic summation** of the potentials contributed by each charge at the point.

$$V_{\text{resultant}} = V_1 + V_2 + V_3 + \dots$$



Remember to keep the signs in your substitution.

⁴ The actual derivation of this equation is not in the syllabus. If you are interested, you can find it in Appendix A.

Example 4

Four charges are placed at the corners of a square as shown in the figure below. Each side has a length a . Find the electric potential at P , the centre of the square.

Potential is given by $\frac{kq}{r}$

For each of the charges, r is equal to $0.5\sqrt{2}a$

Hence,

$$V_1 = \frac{kQ}{0.5\sqrt{2}a}$$

$$V_2 = \frac{2kQ}{0.5\sqrt{2}a}$$

$$V_3 = \frac{2kQ}{0.5\sqrt{2}a}$$

$$V_4 = \frac{-2kQ}{0.5\sqrt{2}a}$$

Adding this all up & plugging in $k = \frac{1}{4\pi\epsilon_0}$,

$$V_{\text{resultant}} = 3\sqrt{2} \frac{Q}{4\pi\epsilon_0 a}$$

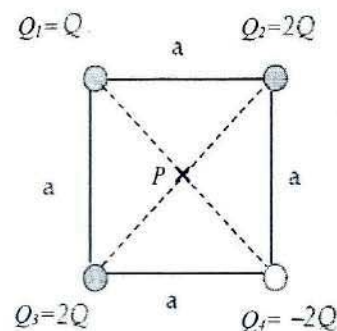


Fig. 14. Figure for Example 4

12.3.2 Electric Potential Energy

The electric potential energy, U , of a charge at a point in an electric field is defined as the work done by an external agent in moving the charge from infinity to that point, without any change in the kinetic energy.

Relationships between U and V :

Electric Potential Energy U of a charge Q placed at a point in an electric field with potential V is given by

$$U = QV$$

The work done by the external agent in moving a charge Q from point A to point B in an electric field is given by

$$U_{A \rightarrow B} = U_B - U_A = Q(V_B - V_A)$$

where V_A and V_B are the potentials of the field at A and B respectively.

Notes:

- (i) U may be positive or negative, depending on the sign of the charge and the potential.
- (ii) Similar to gravitational potential energy, the electric potential energy is stored in a system of charges and not possessed by a single charge Q .

The electrical field is a conservative field which means that

⇒ work done on a charged particle in moving it from one point to another depends only on the particle's initial and final positions and is **independent on the path taken**.

⇒ **the change in potential energy is also independent of the path taken.**

- For example, suppose there are two ways to move a charge Q from point A to point B in an electric field (Path X or Path Y).

Since the field is conservative,

Work done by an external force in bringing a charge Q from A to B via Path X

= Work done by an external force in bringing a charge Q from A to B via Path Y

By conservation of energy (assuming no changes in kinetic energy), the change in potential energy of the system when Q is moved from A to B,

$$\Delta U_{A \rightarrow B} = U_B - U_A = Q(V_B - V_A)$$

is independent of whichever path is taken.

- Furthermore, if the charge Q is brought from A to B through path X and then from B to A through path Y, then the net change in potential energy of the system

$\Delta U_{A \rightarrow B \rightarrow A} =$ work done in bringing the charge from A to B back to A is 0

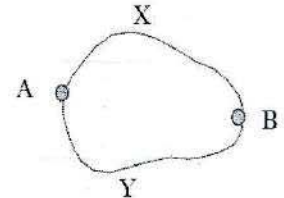


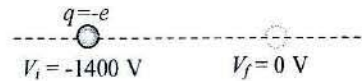
Fig 15

Example 5

An electron in the vacuum of a cathode ray tube moves from rest at a point where the potential is -1400 V to a point where the potential is zero.

Calculate

- the change in its potential energy,
- the change in its kinetic energy, and
- its final velocity.



a) $EPE = qV$

$$EPE_i = qV_i = (-1.6 \times 10^{-19} \text{ C}) \times (-1400 \text{ J C}^{-1})$$

$$EPE_f = qV_f = (-1.6 \times 10^{-19} \text{ C}) \times 0 = 0$$

$$\Delta EPE = -2.24 \times 10^{-16} \text{ J}$$

b) From PCOE, $\Delta KE = -\Delta EPE = 2.24 \times 10^{-16} \text{ J}$

c) Since $KE = \frac{1}{2}mv^2 = 2.24 \times 10^{-16} \text{ J}$

$$v = \left(\frac{4.48 \times 10^{-16} \text{ J}}{9.11 \times 10^{-31} \text{ kg}} \right)^{\frac{1}{2}}$$

$$= 2.2 \times 10^7 \text{ m/s}$$

Potential Energy of a system of Two Point Charges

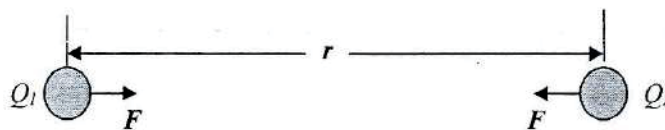


Fig. 16. Two charges Q_1 and Q_2 are separated by a distance r . What is the potential energy of this system?

Consider two point charges Q_1 and Q_2 separated by a distance r (Fig. 16). We can find the potential energy by viewing the situation as such: Q_1 sets up an electric field around itself and Q_2 is in the electric field of Q_1 .⁵ Hence,

$$\text{Potential of the electric field by } Q_1 \text{ at the location of } Q_2, V = \frac{Q_1}{4\pi\epsilon_0 r}$$

$$\text{Potential energy of the system} = Q_2 V = \frac{Q_1 Q_2}{4\pi\epsilon_0 r}$$

Potential Energy Between Two Point Charges:

The potential energy U between two point charges Q_1 and Q_2 separated by a distance r is given by

$$U = \frac{Q_1 Q_2}{4\pi\epsilon_0 r}$$



In Coulomb's Law, as we are looking at the electrical force between the charges, any **negative sign represents the direction** and since we only wish to find the magnitude we ignore the signs of the charges.

Potential energy however is a **scalar quantity**, and the sign no longer indicates the direction. Hence, we **have to indicate the sign in the calculation!**

What is the potential energy of a system of two or more charges?

Consider the system of point charges given by the diagram below. Let us determine the expression for the potential energy of this system.

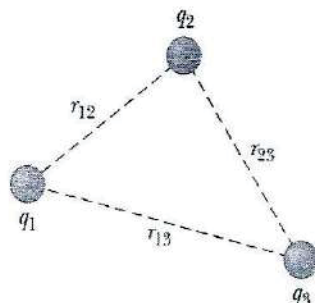


Fig. 17. Potential Energy of the System

⁵ You can also analyse the situation by treating Q_2 as setting up the field and Q_1 being in the electric field by Q_2 . You would still end up with the same final expression for the electric potential energy for the system of two point charges.

To determine the potential energy of this system, we go back to the fundamentals of what potential energy is – the work done by an external force to bring a charge from infinity to the point in the electric field.

So we start off assuming that all the charges are at infinity and infinitely far away from each other. To determine the potential energy of the system, we carry out the following steps and determine the net work required to assemble the whole system.

Step One: Bring q_1 from infinity to the point. As there is no electric field initially in the region,

$$\Rightarrow \text{Work done, } W_1 = 0 \text{ J}$$

Step Two: Bring q_2 from infinity to the point. As q_1 is already in place, it will have set up an electric field in the region and an external force will need to do work against the electrical force experienced by q_2 due to its interaction with the field set up by q_1 . Hence,

$$\Rightarrow \text{Work done, } W_2 = \frac{q_1 q_2}{4\pi\epsilon_0 r_{12}}$$

Step Three: We bring q_3 from infinity to its position in the electric field. Since q_1 and q_2 are now in place, the electric field in the region is now due to q_1 and q_2 , and the external force will need to do work against the force on q_3 due to its interaction with the field set up by both q_2 and q_3 . Hence, the work done in this process is

$$\Rightarrow \text{Work done, } W_3 = \frac{q_1 q_3}{4\pi\epsilon_0 r_{13}} + \frac{q_2 q_3}{4\pi\epsilon_0 r_{23}}$$

Hence, the net potential energy of the system is the net work to assemble the system:

$$\begin{aligned} \Rightarrow \text{Potential energy of this system, } U &= W_1 + W_2 + W_3 \\ &= \frac{q_1 q_2}{4\pi\epsilon_0 r_{12}} + \frac{q_1 q_3}{4\pi\epsilon_0 r_{13}} + \frac{q_2 q_3}{4\pi\epsilon_0 r_{23}} \end{aligned}$$

Potential energy of a System of More than Two Point Charges

Hence, for a system that contains more than two point charges, the

Total potential energy can be found by calculating U for every pair of charges and summing the terms algebraically.

12.4 Relation between Potential Energy and Electrical Force

In the chapter of "Work, Energy and Power", we learnt that the relationship between this force F and the potential energy of the charge U can be written as⁶

$$F = -\frac{dU}{dr}$$

Note:

- This equation gives us both the magnitude and direction of the force acting on the charge placed at a point in an electric field.
 - The *magnitude* of the force is given by:

$$F = \left| \frac{dU}{dr} \right| = |\text{gradient of the } U \text{ vs. } r \text{ graph}|$$

- The direction of the force on the charge is given by the negative (-) sign i.e. **the force on the charge points in the direction where the decrease in potential energy is greatest.**⁷

12.4.1 Relation between Potential Gradient and Electric Field Strength

Now, from above, we have

$$F = -\frac{dU}{dr}$$

$$\Rightarrow qE = -\frac{d(qV)}{dr}$$

where q is the charge placed the point in the electric field with field strength E and potential V . Dividing both sides by q , we get the relationship

$$E = -\frac{dV}{dr}$$

Note:

- $\frac{dV}{dr}$ is known as the **potential gradient** of the field. It gives the variation of the potential of field at the point of consideration.
- The *magnitude* of E is given by :

$$E = \left| \frac{dV}{dr} \right| = |\text{gradient of the } V \text{ vs. } r \text{ graph}|$$
- The *direction* of E is **towards decreasing potential**.

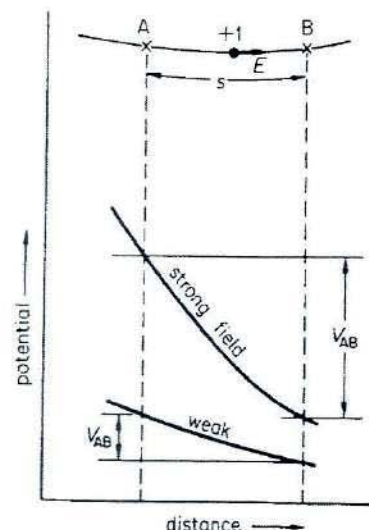


Fig. 18. Relationship between potential and field strength.

⁶ The derivation of $F = -\frac{dU}{dr}$ can be found in the chapter on "Work, Energy and Power" or, alternatively, in Appendix B.

⁷ In fact this is something fundamental in physics, for many systems, to achieve greater stability charges would try to move to a location such that the system will end up with lower potential energy.

12.5 Representation of Electric Fields by Equipotential Lines or Surfaces

Another way to represent the electric field in a region is through *lines* or *surfaces* of equal potential.

An equipotential surface is a surface in which every point on the surface is at the same potential.

It is standard practice to draw these equipotential lines or surfaces the same number of volts apart. Recall, that the relationship between *potential* V and *electric field strength* E is given by $E = -\frac{dV}{dr}$. Hence, from the equipotential lines or surfaces, the electric field can be "visualised".

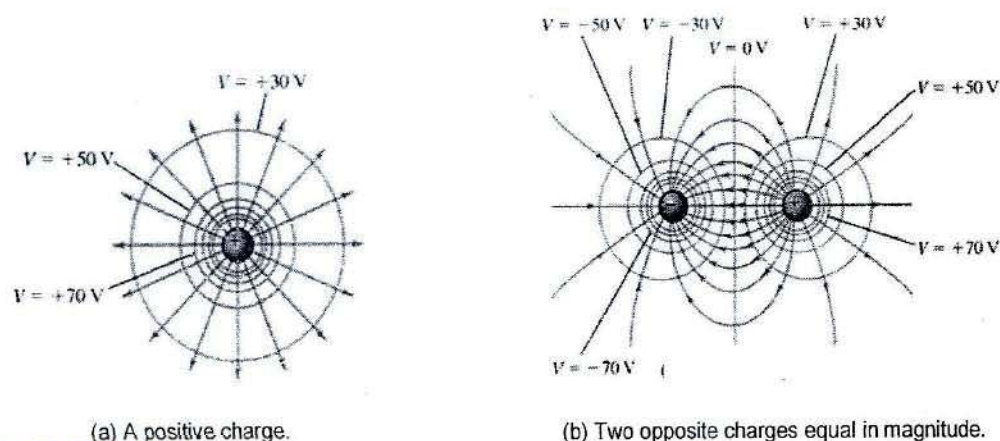
In general:

- Magnitude of the electric field $E = \left| \frac{dV}{dr} \right|$. Hence, **the closer the equipotential lines or surfaces are to each other, the greater the change in potential for the same distance and hence the stronger the electric field.**
- Direction of the **electric field points from higher potential to lower potential.**

Further, all points on an equipotential surface have the same potential, hence

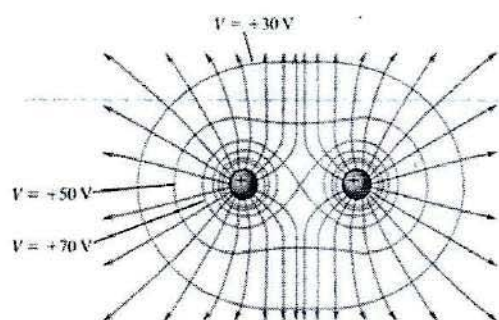
- ⇒ Potential gradient along an equipotential surface is zero.
- ⇒ There is no component of electric field along an equipotential surface.
- ⇒ **Electric field lines are always perpendicular to equipotential surfaces.**

Fig. 19 below shows some examples of equipotential surfaces for different systems.

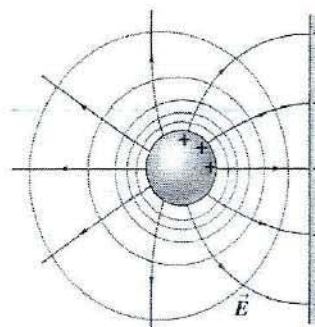


(a) A positive charge.

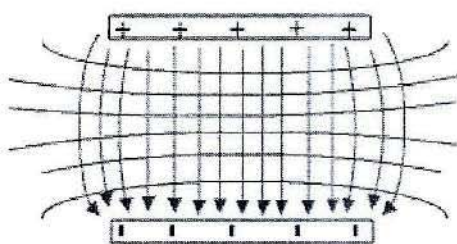
(b) Two opposite charges equal in magnitude.



(c) Two equal positive charges.



(d) A positively charged sphere and a negatively charged plate.



(e) Two parallel plates with opposite charge.

→ electric field lines — equipotential lines (cross-section of equipotential surfaces)

Fig. 19. Equipotential and field lines.

Note : In such diagrams, only some of the equipotential lines are drawn, there could be more equipotential lines nearer or further from the source apart from those shown in the diagram.

Charges in Equilibrium on Conductors

A good electrical conductor contains charges that are free to move. Hence, when a conductor is charged, the charges will always redistribute themselves until the excess charge resides on the surface (they try to move out of the conductor as they repel each other) and till the net force on them is zero. When there is no net motion, the conductor is said to be in **electrostatic equilibrium**.

Since the charges at the surface experience no force parallel to the surface (else they will continue to move along the surface),

From $F_{\parallel} = qE_{\parallel} = 0$

⇒ **Electric field along the surface is zero, $E_{\parallel} = 0$.**

From $E = -\frac{dV}{dr}$

⇒ $\frac{dV}{dr} = 0$ along the surface of the conductor

⇒ **The surface of a conductor is an equipotential surface.**

⇒ **Electric field lines emerging or ending on a conductor are perpendicular to the surface.**

Furthermore, using calculus, it is possible to show that the electric field within a conductor is zero, hence there is no potential gradient within the conductor and

the potential in the conductor is constant everywhere inside the conductor and equal to its value at the surface.

⇒ **The electric field in the conductor is therefore zero (no potential gradient). Hence, no field lines are drawn.**

With this we can now summarise a more complete set of rules for drawing electric field lines and equipotential lines for charged systems. We will include the previous rules for completeness.

Rules for Drawing Electric Field Lines and Equipotential lines:

- i. Electric field lines must begin from a positive charge and end on a negative charge. In the case of an excess of one type of charge, some lines will begin or end infinitely far away.
- ii. The number of lines drawn leaving a positive charge or approaching a negative charge is proportional to the magnitude of the charge.
- iii. Field lines do not intersect, because E at a point can only point in one direction.
- iv. Equipotential lines/surfaces are lines/surfaces that join all points of equal potential.
- v. Equipotential lines/surfaces are drawn at equal intervals of potentials (e.g. 10 V, 20 V, 30 V etc.). Hence, the closer the equipotential lines or surfaces, the stronger the field.
- vi. Electric field lines are always perpendicular to lines of equal potential, e.g. to the surface of a conductor.
- vii. Electric field lines always point towards the direction in which potential is decreasing.
- viii. The electric field in a conductor is zero. No field lines are drawn.

12.6 Electric Field between Two Charged Parallel Plates

Consider two large metal plates that carry equal but opposite charges. (This can be done by connecting them to the opposite terminals of a constant voltage supply.) It is reasonable to suppose that within each plate, except at the edges, the charges on each plate are spread uniformly over the inside surface of each plate, because of their mutual repulsion and the attraction by the opposite charges on the other plate.

The lines of forces are straight, parallel to each other and equally spaced (except at the fringes – see Fig. 8(f)). **The electric field in the space between the two plates is said to be uniform.**

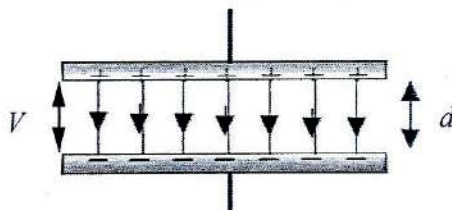


Fig. 20. Electric Field between two parallel, oppositely charged plates.

Suppose that in Fig. 20, the potential difference between two large charged metal plates is V and the plate separation is d . Since the electric field strength between the plates is constant, from $E = -\frac{dV}{dr}$ we have

- **Magnitude of the electric field strength:**

$$E = \frac{V}{d}$$

- **Direction of the electric field:** E is directed from the plate with the higher potential to the plate with the lower potential.

12.6.1 Motion of Charged Particles in a Uniform Electric Field

Consider a particle of mass m , carrying charge q , that is placed in a uniform electric field E . The field exerts a force qE on it, giving it an acceleration a in the direction of the force.

By Newton's second law, $ma = qE$. The acceleration of the charge is therefore given by $a = \frac{qE}{m}$.

In most questions, the mass is so small that the electric force overwhelms the gravitational force on the mass (weight). Hence the weight can be ignored.



For example, if you are considering an electron:

Its weight is $9.11 \times 10^{-31} \times 9.81 = 8.94 \times 10^{-30} \text{ N}$

If it is between parallel plates of potential difference 1.0 V and separation 10 cm, the electric force it experiences is $(1.60 \times 10^{-19}) \times 1.0/0.10 = 1.6 \times 10^{-18} \text{ N}$.

However, you should not assume that it is always the case that the weight can be ignored. You might get a question where the mass is large enough to make the weight significant compared to the electric force. (e.g. the Millikan oil drop experiment).

A positively charged particle will accelerate in the direction of E . A negatively charged particle will accelerate in the direction opposite to E .

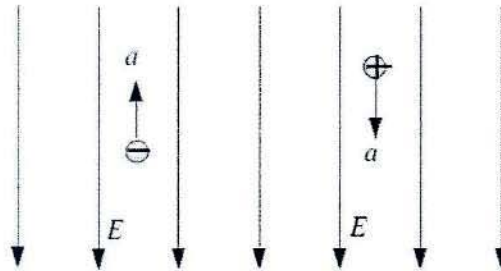


Fig. 21. Acceleration of charged particles subjected to an E-field.

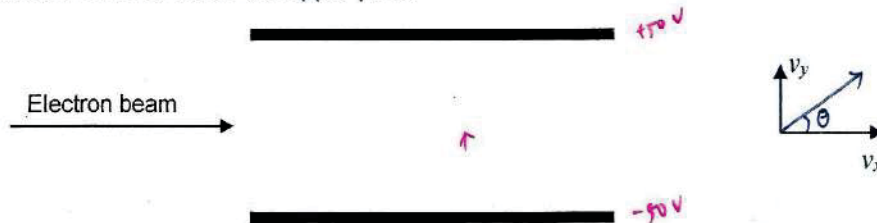
In general, the motion is that of an object moving under constant acceleration. There are two possible paths:

- i. If its initial velocity is zero or parallel to the direction of E , the charged particle will move in a straight line parallel to E .
- ii. If its initial velocity is inclined to the field E , it will travel in a parabolic path.



Example 6

A beam of electrons, travelling at $1.35 \times 10^7 \text{ m s}^{-1}$, enters a uniform electric field between two plates of length 0.060 m , which are separated by a distance of 0.020 m . The top plate is at $+50 \text{ V}$ while the lower plate is at -50 V . Find the angular deflection of the beam as electrons emerge from between the plates, assuming that the electrons never reach the upper plate.



No horizontal forces

$$v_x = u_x = 1.35 \times 10^7 \text{ m/s}$$

Vertically, we have $F_e = eE = e \left(\frac{\Delta V}{d} \right)$ upwards

Plugging in $e = 1.60 \times 10^{-19} \text{ C}$

$$\Delta V = 100 \text{ V}, d = 0.020 \text{ m}, m_e = 9.11 \times 10^{-31} \text{ kg}$$

$$\text{we get } a_y = F_e / m_e = 8.781 \times 10^{14} \text{ m/s}^2$$

$$a_y = \frac{F_e}{m_e}$$

$$\text{Time of flight is given by } t = \frac{s_x}{v_x} = 4.444 \times 10^{-9} \text{ s}$$

Using kinematics,

$$v_y = u_y + a_y t = a_y t$$

$$= 3.902 \times 10^6 \text{ m/s}$$

$$\tan \theta = \frac{v_y}{v_x}$$

$$= 16.1^\circ$$

12.7 Analogy between Electric fields and Gravitational Fields

The laws governing interaction between point charges and the interaction between point masses are analogous. Mass, m , is analogous to charge, Q . The constant G is analogous to $1/4\pi\epsilon_0$.

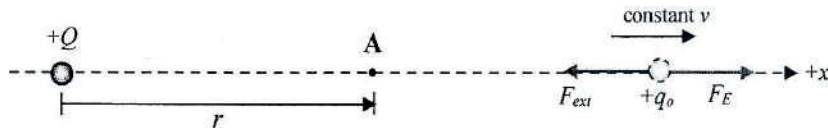
	Electric field	Gravitational field
Field Strength	$\vec{E} = \frac{\vec{F}}{q}$	$\vec{g} = \frac{\vec{F}}{m}$
	Point charge: $E = \frac{Q}{4\pi\epsilon_0 r^2}$	Point mass: $g = \frac{GM}{r^2}$
Force	Coulomb's Law: $F = \frac{1}{4\pi\epsilon_0} \frac{Qq}{r^2}$ The force between point charges follows an inverse square law, i.e. $F \propto \frac{1}{r^2}$	Newton's Law of Gravitation: $F = G \frac{Mm}{r^2}$ The force between point masses also follows an inverse square law, i.e. $F \propto \frac{1}{r^2}$
	Force is a <i>vector</i> quantity. The resultant force at a point must be determined by <i>vector sum</i> of the forces.	
Potential	Due to a point charge: $V = \frac{Q}{4\pi\epsilon_0 r}$ Potential can be positive or negative depending on whether Q is positive or negative.	Due to a point mass: $\phi = -\frac{GM}{r}$ Potential is always negative.
	Potential is a <i>scalar</i> quantity. The resultant potential at a point is determined by the <i>algebraic sum</i> of the potential due to each charge or mass.	
Potential Energy	$U = Vq$ Between point charges: $U = \frac{Qq}{4\pi\epsilon_0 r}$ U can be positive or negative. It is positive if Q and q have the same sign. Otherwise, U is negative.	$U = m\phi$ Between point masses: $U = -\frac{GMm}{r}$ U is always negative.
Field Strength and potential gradient	$E = -\frac{dV}{dr}$ The negative sign indicates the <u>direction</u> of the field is in the direction of decreasing potential.	$g = -\frac{d\phi}{dr}$ The negative sign indicates the <u>direction</u> of the field is in the direction of decreasing potential.
Force and potential energy gradient	$F = -\frac{dU}{dr}$	$F = -\frac{dU}{dr}$

Appendices

Appendix A: Deriving the Potential of an Electric Field by a Point Charge

[Adapted from: Duncan, 2000, p 204]

Consider a point charge $+Q$ placed at a point O. To find the potential of the electric field at A a distance r away from $+Q$, we bring a very small test charge $+q_0$ from infinity to the point A (allowing for no gain in kinetic energy) and find the work done by the external force to carry out this action.



We note that when the test charge is moving in the field set up by the source, it experiences a varying electrical force F_E . An external force needs to move it from infinity to the point.

Now, the work done δW by the external force to move a small distance of δx towards the source charge is

$$dW = F_{ext} dx$$

Since the test charge is moved at a constant velocity, $F_{ext} = -F_E$. Therefore,

$$dW = -F_E dx$$

The electric force experienced by the test charge a distance x away from the source charge is

$F_E = \frac{Qq_0}{4\pi\epsilon_0 x^2}$. Therefore, we have the net work done by the external force,

$$W = -\int_{\infty}^r F_E dx = -\int_{\infty}^r \frac{Qq_0}{4\pi\epsilon_0 x^2} dx = -\frac{Qq_0}{4\pi\epsilon_0} \left[-\frac{1}{x} \right] = \frac{Qq_0}{4\pi\epsilon_0 r}$$

By definition of the electrical potential, the potential at point A is therefore,

$$V = \frac{W}{q_0} = \frac{Q}{4\pi\epsilon_0 r}$$



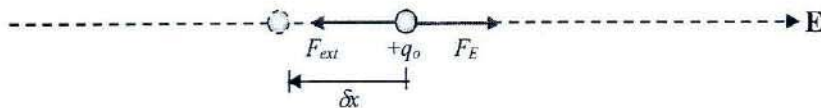
Appendix B: Deriving the Relationship between Force and Potential Energy

The relationship between a (conservative) force F and the associated potential energy is given by

$$F = -\frac{dU}{dx}$$

This is a general relationship that applies to all conservative forces, of which the gravitational force and the electric force are examples.

To get this relationship, it is best to look at a specific case. In this case we will look at the electrical force F_E experienced by a charge in an electrical field.



By conservation of (mechanical) energy, assuming that no dissipative forces act on the system,

Work done by an external force on the system W = Increase in P.E. + Increase in K.E. of system

Assuming we move the object in the field at a constant velocity

$\Rightarrow$ Increase in K.E. of system = 0

Now, let us move the object (a small charge in this case) by a small distance of δx in the direction of the external force, the small amount of work done δW by the external force in this case can be written as

$$\delta W = F_{ext} \delta x$$

Since the object is moving with constant velocity, $F_{ext} = -F_E$

Hence, we have

$$\delta W = F_{ext} \delta x$$

$$\Rightarrow \delta W = -F_E \delta x$$

Now, as stated above, the work done by the external force results in the gain of potential energy of the system, hence we have

$$\Rightarrow \delta U = \delta W = -F_E \delta x \quad \text{where } U \text{ is the potential energy}$$

which can be written as $\Rightarrow F_E = -\frac{dU}{dx}$ (derived)

Note: In our derivation, the scenario given is that of a charge moved in an electric field but there was no substitution of the actual force or field. The scenario is given to only help in the visualisation. The same derivation can be used for the scenario of a mass moved in a gravitational field or even a spring stretched by an extension of x ; you would still be able to get the same general result.

Topic 13: Current of Electricity & D.C. Circuits

C2 2017

H2 Physics (Syllabus 9749)

Current of Electricity

Content

- Electric Current
- Potential Difference
- Resistance and resistivity
- Electromotive force

Learning Outcomes

Candidates should be able to:

- show an understanding that electric current is the rate of flow of charge
- derive and use the equation $I = nAvq$ for a current-carrying conductor, where n is the number density of charge carriers and v is the drift velocity
- recall and solve problems using the equation $Q = It$
- recall and solve problems using the equation $V = W/Q$
- recall and solve problems using the equations $P = VI$, $P = I^2R$ and $P = V^2/R$
- recall and solve problems using the equation $V = IR$
- sketch and explain the $I - V$ characteristics of various electrical components such as an ohmic resistor, a semiconductor diode, a filament lamp and a negative temperature coefficient (NTC) thermistor.
- sketch the resistance- temperature characteristic of an NTC thermistor.
- recall and solve problems using the equation $R = \frac{\rho l}{A}$
- distinguish between electromotive force (e.m.f.) and potential difference (p.d.) using energy considerations.
- Show an understanding of the effects of the internal resistance of a source of e.m.f. on the terminal potential difference and output power.

D.C. Circuits**Content:**

- Circuit symbols and diagrams
- Series and parallel arrangements
- Potential Divider
- Balanced potentials

Learning Outcomes:

Candidates should be able to:

- recall and use appropriate circuit symbols as set out in the ASE publication *Signs, Symbols and Systematics* (*The ASE Companion to 16-19 Science, 2000*)
- draw and interpret circuit diagrams containing sources, switches, resistors, ammeters, voltmeters and/or any other type of component referred to in the syllabus.
- solve problems using the formula for the combined resistance of two or more resistors in series.
- solve problems using the formula for the combined resistance of two or more resistors in parallel.
- solve problems involving series and parallel circuits for one source of e.m.f.
- show an understanding of the use of a potential divider circuit as a source of variable p.d..
- explain the use of thermistors and light-dependent resistors in potential divider circuits to provide a potential difference which is dependent on temperature and illumination respectively.
- recall and solve problems using the principle of the potentiometer as a means of comparing potential differences.

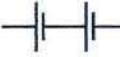
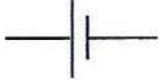
Table of Contents

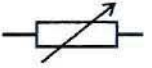
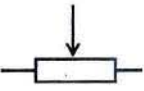
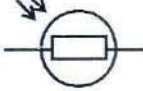
13.1 Practical Circuits.....	5
13.2 Electric Current.....	6
13.2.1 Direction of Electric Current	6
13.2.2 The velocities of the current carriers	7
13.2.3 Units of Current & Charge.....	8
13.3 Potential Difference and Electromotive Force.....	10
13.3.1 Energy considerations of an electric current.....	10
13.3.2 Electromotive Force (e.m.f.).....	10
13.3.3 Potential Difference.....	11
13.3.4 Distinction between e.m.f. and p.d.	12
13.4 Electrical Power.....	12
13.5 Resistance and Resistivity.....	14
13.5.1 Resistance.....	14
13.5.2 Ohm's Law (Review of High School Physics).....	14
13.5.3 Resistivity.....	15
13.5.4 Determination of resistance using an ammeter and a voltmeter.....	16
13.5.5 Effective Resistance for Combination of Resistors.....	17
13.5.6 Power dissipation in resistors (Joule heating)	19
13.6 I-V Characteristics of Various Electrical Components.....	21
Metallic Conductor	21
Filament Lamp (case of a metal conductor whose temperature is not controlled).....	21
Thermistor.....	22
13.7 Internal resistance of a source of e.m.f.....	24
13.7.1 Terminal potential difference, e.m.f and internal resistance.....	24
13.8 The Potential Divider Principle.....	26
13.8.1 Use of thermistors and light dependent resistors in potential dividers	27
13.9 The Potentiometer.....	29
13.9.1 Applications of the potentiometer.....	29
APPENDIX A – Two different arrangements of voltmeter and ammeter to find a more accurate value of the resistance of an unknown resistor.	34
APPENDIX B – The Wheatstone bridge (Good to know).....	35

13.1 Practical Circuits

Symbols for various circuit elements

Conductor		Ammeter	
Junction of conductors		Voltmeter	
Earth		Galvanometer	

Switch	 Open  Closed	Battery (combination of cells)	 or 
Cell		Lamp	

Fixed resistor		Diode	
Variable resistor		Thermistor	
Potential divider		Light dependent resistor	

Circuit Essentials

Most complete electric circuits contain 6 parts:

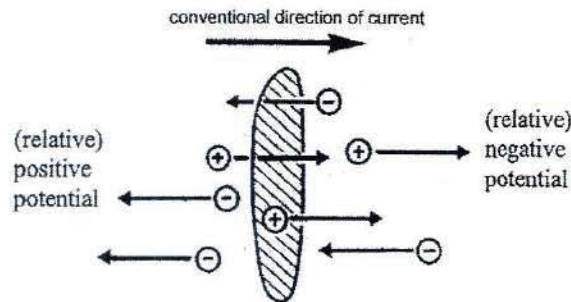
1. An *energy source* to provide the voltage needed to drive current through the circuit.
2. *Conductors* through which the current can travel.
3. *Insulators* to confine the current to the desired paths.
4. A *load* to control the amount of current and convert the electrical energy supplied by the energy source.
5. A *control device*, often a switch, to start and stop the flow of current.
6. A *protection device* to interrupt the circuit in case of a circuit malfunction

The first 4 of the above 6 parts are essential parts. All complete circuits use them. The control device is occasionally omitted. Protection devices are often omitted from circuits.

13.2 Electric Current

Electric current is the rate of flow of electric charge. It can be due to the flow of positive charges or negative charges or both.

Consider some charges moving through a cross-sectional area, which is perpendicular to the flow of charges. Suppose the electric charge δQ flows through the area in time δt .



The electric current I through the area is the rate of flow of the electric charges through the cross-sectional area of the surface. $I = \frac{\delta Q}{\delta t}$

If the charge flows at an irregular or non-linear rate, then the current I_i due to this flow of charge is

$$I_i = \frac{dQ}{dt} \quad (\text{Formula for instantaneous current})$$

If the charges flow such that in a time interval t , the total charge which crosses that area is Q , the average current I_{av} due to this flow of charge is

$$I_{av} = \frac{Q}{t} \quad (\text{Formula for average current})$$

13.2.1 Direction of Electric Current

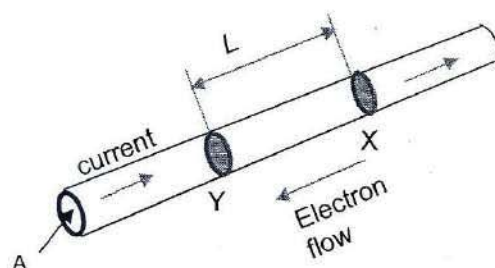
By convention, the direction of electric current (also known as conventional current) is chosen to be in the **same direction as the flow of positive charges**.



If the conductor is a metal (e.g. copper), the current is due to the motion of negatively charged electrons (i.e. mobile valence electrons). By convention, the current due to the flow of negative charges to the left for example is treated as a current towards the right. **Conventional current is in the opposite direction to the flow of electrons.**¹

¹ There is an interesting story on why the direction of current is always interpreted based on the flow of positive charges and not that of negative charges. Do find out more by investigating the work of Benjamin Franklin (1706 - 90) who was the first person to study the electric current.

13.2.2 The velocities of the current carriers



The diagram above shows a length of conductor of cross sectional area $A \text{ m}^2$ containing n free (mobile) charges per m^3 (also known as the number density of charge carriers). The number of free charges (or electrons) in the conductor has to remain constant; when a current flows in a metallic wire, electrons enter it at one end and the same number leave at the other per unit time.

The motion of the electrons is not rectilinear but rather somewhat irregular, but we can assign an average or drift velocity $v \text{ m s}^{-1}$ to the flow along the wire.

Suppose the section of the conductor between X and Y has a length L and a cross-sectional area A , the total number N of charge carriers in the volume between X and Y is

$$N = nAL.$$

If each charge carrier has a charge of q , then the total charge passing the point Y in time Δt is Nq .

Therefore the current I passing point Y is

$$\begin{aligned} I &= \frac{\text{total charge passing Y}}{\text{time taken}} = \frac{Nq}{\Delta t} \\ &= \frac{(nAL)q}{\Delta t} \\ &= nA \left(\frac{L}{\Delta t} \right) q \end{aligned}$$

$$I = nAvq$$

For a given current I , a good conducting material is one for which n is large and v therefore small. Copper is one of the best conductors, having in the region of 10^{29} free electrons per m^3 .

Using the equation above, we can estimate a typical value for v when a current of 1A flows through a copper wire of cross sectional area $2.0 \times 10^{-7} \text{ m}^2$ (about 0.5 mm diameter).

$$\begin{aligned} v &= \frac{I}{nAe} \\ &= \frac{1}{10^{29} \times 2 \times 10^{-7} \times 1.6 \times 10^{-19}} \\ &= 3.1 \times 10^{-4} \text{ m s}^{-1} \end{aligned}$$

KIASU
ExamPaper

The actual values of the velocity v are amazingly small! However it is worth noting that this slow drift starts almost instantaneously in all parts of an electrical circuit completed when, for instance, a switch is closed. (Hence there is clearly no delay between closing a switch in one part of a room and the light in another part coming on.)

13.2.3 Units of Current & Charge

Current

SI unit for electric current is **ampere (A)**, named after the French scientist who established the laws of electromagnetism. It is defined in terms of the magnetic forces that act between current-carrying conductors. The definition will be given in the topic Electromagnetism.

Charge

SI unit for charge is **coulomb (C)**, after the French scientist who first investigated the forces acting between electric charges. The unit is defined in terms of SI base units, ampere (A) and second (s).

One **coulomb (C)** is the quantity of electric charge that pass a given point in a circuit when a steady current of one ampere flows through that point for one second.

$$1 \text{ coulomb} = 1 \text{ ampere-second}$$

$$1 \text{ C} = 1 \text{ A s}$$

If a steady current I flows in a conductor for a time interval t , the quantity of electric charge Q that passes a cross-section of the conductor is given by the product of the current and time.

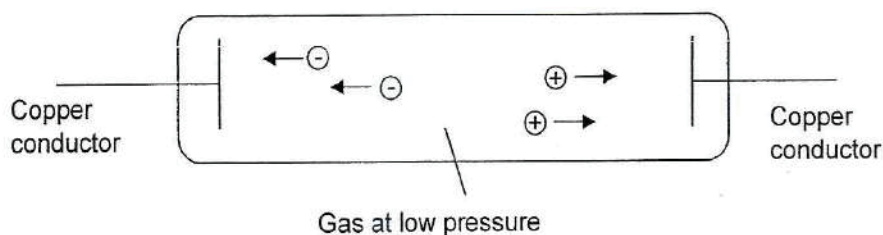
$$Q = It$$

The magnitude of the charge on an electron is e , i.e. $1.60 \times 10^{-19} \text{ C}$.

The total charge Q of a quantity N of electrons is

$$Q = Ne$$

However, it is difficult to count the exceedingly large number of charge particles that constitute a measurable or observable steady current. Therefore, the unit for charge is not defined in terms of e , but in terms of current I .

Example 1

An external electric field is applied to the gas in the tube, resulting in negative particles flowing in one way and positive particles flowing in the opposite direction as illustrated above. In 1 s, 1.56×10^{15} negative particles pass through a perpendicular cross-section in the tube and the same amount of positive particles pass through it in the opposite direction. If each positive particle carries a charge $+2e$, and each negative particle carries a charge $-e$,

- (a) Find the current due to the flow of positive charges.

$$I_{\text{positive}} = Ne(2) \\ = (1.56 \times 10^{15})(1.6 \times 10^{-19})(2) = 4.99 \times 10^{-4} \text{ A}$$

- (b) Find the current due to the flow of negative charges.

$$I_{\text{negative}} = (1.56 \times 10^{15})(1.6 \times 10^{-19}) = 2.50 \times 10^{-4} \text{ A}$$

- (c) Find the net current through the gas.

$$I_{\text{net}} = I_{\text{positive}} + I_{\text{negative}} = 7.49 \times 10^{-4} \text{ A}$$

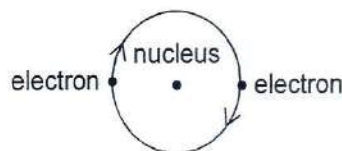
Example 2

The diagram shows a model of an atom in which two electrons move round a nucleus in a circular orbit. The electrons complete one full orbit in $1.0 \times 10^{-15} \text{ s}$.

What is the current caused by the motion of the electrons in the orbit? (J99/I/5)

$$I = \frac{Q}{t} = \frac{Ne}{t} \\ = \frac{2 \times (1.60 \times 10^{-19})}{1.0 \times 10^{-15}}$$

$$= 3.2 \times 10^{-4} \text{ A}$$



13.3 Potential Difference and Electromotive Force

13.3.1 Energy considerations of an electric current

The circulation of an electric current in a complete circuit leads to the continual dissipation of energy throughout the circuit in the form of heat and other forms of energy (e.g. light energy in a filament lamp, mechanical energy in an electric motor, etc). All this energy must be continuously supplied by a voltage source (or more correctly, an e.m.f.), e.g., battery, dynamo, etc.

13.3.2 Electromotive Force (e.m.f.)

The sources of e.m.f. (stands for **electromotive force**) required in any complete circuit to generate electric current are therefore primarily means of **converting energy to electrical form from some other form**. E.g. a dry cell converts chemical energy to electrical energy, a dynamo converts kinetic energy to electrical energy, a microphone converts vibrational energy of a sound wave to electrical energy, etc.



The term e.m.f. has historical origins, but it is **NOT a force**, despite its name. It is more a measure of energy (i.e., electrical potential energy) rather than force.

The **electromotive force (e.m.f.)**, $\mathcal{E}$ of a source is defined as the energy converted per unit electric charge to electrical energy in driving charge round a complete circuit.

The SI unit for e.m.f. is **volt (V)**, after the Italian scientist Alessandro Volta, who first investigated the action of electric batteries.

If charge Q flows in a circuit containing a source of e.m.f. $\mathcal{E}$, the electrical energy supplied by the source to the whole circuit is

$$W = Q\mathcal{E}$$

The power P of the source varies with the current it drives and hence, we normally specify the power of the source with reference to the current, i.e. its power to current ratio. This ratio which is a constant of the electrical source is the e.m.f. of the source.

$$P = \frac{dW}{dt} = \frac{dQ}{dt} \mathcal{E} = I\mathcal{E}$$

$$\mathcal{E} = \frac{P}{I}$$

13.3.3 Potential Difference

A source of e.m.f. produces an *electric field* between its terminals. The electric force will act on charge carriers in the circuit and a complete conducting path between the terminals allows the charge carriers to flow through the circuit.

Inside the conductors, the electric force does work on the moving charged particles. As a result, electrical potential energy is converted into heat and other forms of energy.

We can define the electric potential at a point in the circuit. The conventional current will flow from a point of higher electric potential to a point of lower electric potential in the circuit. The potential difference (p.d.), between two points in the circuit is the work done by the electric field to drive a unit charge from one point to the other.

The p.d. across a conductor is sometimes described as "voltage" or "potential drop" across the conductor.



Most of the time, ***it is only differences of potential (i.e., p.d.) and NOT the potentials of points in a circuit that are of any significance in electric circuits.***

To determine the electric potential at a point in a circuit, it would be necessary to first identify a point to be at conventional zero potential. This point is used as a reference to which the potentials of other points in the circuit can be determined. Many electrical circuits have a point in the circuit that is connected to Earth. By convention, *Earth is taken to be at zero potential*. Electrical apparatus that is connected to the earth is said to be "earthed".



Circuit symbol to represent earthing of a point of a circuit (i.e. the point is at 0 V potential).

The **potential difference V between two points** in a circuit is defined as the energy converted per unit electric charge from electrical energy to other forms of energy when charge moved between the two points.

If the electrical energy W converted is for charge Q moved between the two points, then, potential difference V between two points is

$$V = \frac{W}{Q}$$

Thus, if a charge Q flows in a part of a circuit across which there is a p.d. V , the *electrical energy supplied* is given by QV .

The SI unit for potential difference is **volt (V)**.

One volt is the potential difference across two points in a circuit in which one joule of electrical energy is converted to other forms of energy when one coulomb of charge is moved between the two points.

$$1 \text{ V} = \frac{1 \text{ J}}{1 \text{ C}}$$

13.3.4 Distinction between e.m.f. and p.d.

Energy conversion

e.m.f. is a measure of the energy (converted **from other forms of energy to electrical**) a **source** can transfer per coulomb of charge when it is driven through the circuit.

p.d. **between two points in a circuit** is a measure of the energy per coulomb dissipated (**from electrical energy to other forms of energy**) when one coulomb passes through the part of the circuit between the two points.

There will be no p.d. across any part of the circuit without an e.m.f. in the circuit.

An e.m.f. exists whether current flows or not; e.g. a left-alone cell has e.m.f. .

13.4 Electrical Power

If a charge Q flows in a part of a circuit across which there is a p.d. V , the electrical energy supplied by the source to that part of the circuit is given by

$$W = QV \quad \dots\dots\dots (1)$$

The flow of current is accompanied by the conversion of electrical energy into other forms of energy. *The rate at which energy is converted from one form to another in a device is called its power.*

Hence, a device of power P operating for time t converts a quantity of energy W , given by

$$W = Pt \quad \dots\dots\dots (2)$$

From (1) & (2)

$$Pt = QV$$

$$P = \frac{Q}{t} V$$

$$P = IV$$

Hence the **electrical power** P converted by a device when there is a potential difference V across it and a current I flowing through it is given by

$$P = IV$$

The unit of power is the **watt (W)**, after the Scottish engineer James Watt who developed the steam engine. Units of power in regular use are the *kilowatt (kW)* and the *megawatt (MW)*.

The equation above gives rise to an alternative definition of potential difference across any electrical device or in electrical circuits.

The potential difference between two points in a circuit or across any electrical device can also be defined as the electrical power delivered per unit current that flows through it.



$$V = \frac{P}{I}$$

One volt is the potential difference between two points of a conductor carrying a current of one ampere when the power dissipated is one watt.

$$1 \text{ V} = \frac{1 \text{ W}}{1 \text{ A}}$$

Useful units of electrical energy

Recall: The SI unit of energy is the joule (J), named after the British scientist James Prescott Joule who established the relationship between heat and mechanical forms of energy.

$$1 \text{ J} = 1 \text{ N m (newton meter)}$$

i) *The kilowatt-hour (kWh)*

The calculation of quantities of electrical energy for commercial purposes is conveniently carried out by means of this unit.

$$\begin{aligned} \text{Since } 1 \text{ kW} &= 10^3 \text{ J s}^{-1} \text{ and } 1 \text{ hour} = 3.6 \times 10^3 \text{ s,} \\ \text{Therefore, } 1 \text{ kWh} &= 3.6 \times 10^6 \text{ J} \end{aligned}$$

ii) *The electron-volt (eV)*

This is a convenient unit for the calculation of particle energies in atomic and nuclear physics (frequently used in the topic of Electric Field and Nuclear Physics). It is the quantity of energy gained by an electron in accelerating through a p.d. of 1 V.

$$\begin{aligned} \text{Since charge of one electron is } -1.60 \times 10^{-19} \text{ C,} \\ 1 \text{ eV} &= 1.60 \times 10^{-19} \text{ J} \end{aligned}$$

Example 3

The current in a device is 1.2 A and the potential difference across it is 12.0 V.

- (i) What is the power supplied to the device?
- (ii) How much electrical energy is supplied to the device in 30 seconds?

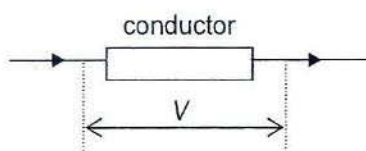
$$\text{(i). } P = IV = 1.2 \times 12.0 = 14.4 \text{ W}$$

$$\text{(ii). } W = Pt = 14.4 \times 30 = 432 \text{ J}$$

13.5 Resistance and Resistivity

If a potential difference is applied across a conductor, then a current I will flow through the conductor. The potential difference required to produce a given current depends on a property of the conductor called its resistance. Resistivity is a property of the material of the conductor. Example: *resistance* of copper wire, *resistivity* of copper.

13.5.1 Resistance



The resistance, R , of a device is the **ratio** of the potential difference V across the device to the current I flowing through it.

$$R = \frac{V}{I}$$

The electrical resistance of a conductor is due to its atoms obstructing the drift of electrons past them. Resistance is thus a measure of the obstruction to the movement of charge carriers.

Unit for resistance R is **ohm (Ω)**.

One ohm is the resistance of a conductor through which a current of one ampere flows when a potential difference of one volt exists across it.

$$1 \Omega = 1 \text{ V A}^{-1}$$

13.5.2 Ohm's Law (Review of High School Physics)

Georg Simon Ohm, a German physicist, was the first person to distinguish clearly between the two concepts, p.d. and current, and to investigate the relation between them. Ohm determined that there is a *direct proportionality* between the potential difference applied across a metallic conductor and the resultant electric current through the conductor (known as Ohm's Law). From the results of his experiment, Ohm was able to define the fundamental relationship between voltage, current and resistance.

The current in a metallic conductor depends on a number of factors beside the p.d. across it. The most important of these is the temperature. But other physical conditions will also affect the current through it.

In addition, for electrolytic conduction, the relation between p.d. and current is complicated by the e.m.f. that arises at the electrodes, accompanying the conversion of electrical energy to chemical energy. The same situation arises in other cases where the conductor is itself the site of an e.m.f. (e.g. the armature coil of a motor, or a coil in which a rapidly changing current induces an e.m.f. by self-induction).

Thus, **Ohm's Law states that:**

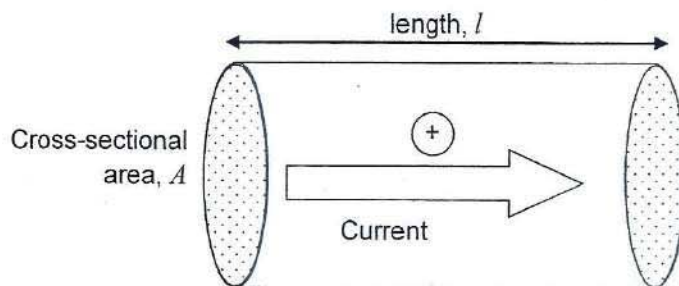
A steady current flowing through a metallic conductor is proportional to the potential difference across it, provided the temperature and other physical conditions are constant.

As a consequence of Ohm's Law, the resistance for a given metallic conductor under steady physical conditions is a constant. Such conductors are also known as **ohmic conductors**.

13.5.3 Resistivity

The resistance of a uniform conductor is

- (i) proportional to its length l
- (ii) inversely proportional to its cross-sectional area A .
- (iii) dependent on a property of the material of the specimen known as its resistivity, ρ .



The relationship between resistance and the above factors is given by

$$R = \rho \frac{l}{A}$$

Unit for resistivity is $\Omega \text{ m}$.

The table shows the resistivity of some common materials.

Material	Resistivity at 20°C / $\Omega \text{ m}$
silver	1.6×10^{-8}
copper	1.7×10^{-8}
gold	2.2×10^{-8}
iron	10.1×10^{-8}
nichrome	110×10^{-8}
skin (dry)	3×10^4

Example 4

The resistivity of the human body is low compared with the resistivity of the skin which is about $3 \times 10^4 \Omega \text{m}$ for dry skin. For a layer of dry skin 1 mm thick, determine the resistance of a 1 cm^2 area of skin.

$$R = \frac{\rho l}{A} = \frac{(3 \times 10^4)(0.001)}{1 \times 10^{-4}} = 3 \times 10^5 \Omega$$

13.5.4 Determination of resistance using an ammeter and a voltmeter

In principle, the resistance R of a resistor can be found by measuring the current I which flows through it as a result of there being a potential difference, V , across it, and substituting the measured values of I and V into the equation $R = \frac{V}{I}$.

Whenever **an ammeter** is put into a circuit, it is put **in series** with the circuit components, and therefore it must have **zero resistance** if it is not to affect the current in the circuit.

A voltmeter, on the other hand, is always placed **in parallel** with a circuit component, and therefore must have **infinite resistance** if it is not to draw current through itself and alter the current in the circuit.

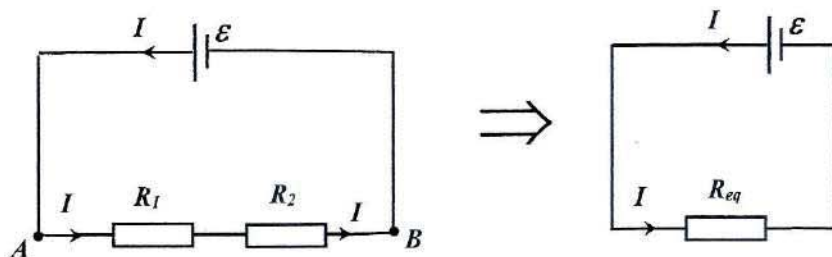
An ideal ammeter has zero resistance; an ideal voltmeter has infinite resistance.

A real ammeter has a resistance which can range from near zero to 1Ω ; a real voltmeter has a resistance, which can range from $20 \text{ k}\Omega$ to $2 \text{ M}\Omega$.

13.5.5 Effective Resistance for Combination of Resistors

Resistors in series

Consider two resistors R_1 and R_2 which are connected in series across A and B.



$$\mathcal{E} = IR_1 + IR_2$$

$$\mathcal{E} = I(R_1 + R_2)$$

$$\frac{\mathcal{E}}{I} = (R_1 + R_2)$$

$$R_{eq} = R_1 + R_2$$

For n resistors connected in series:

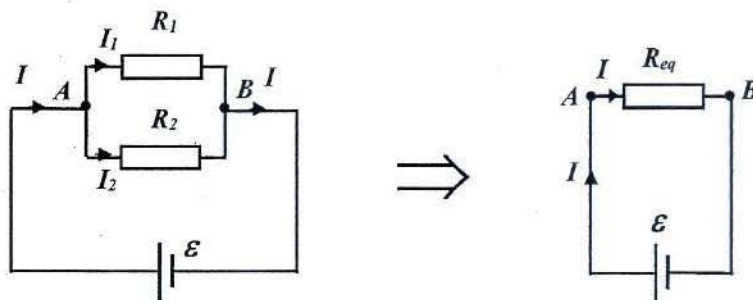
$$R_{eq} = \sum R_n$$

Characteristics of circuits consisting of **resistors in series** are;

- (i) the **same current** flows through each of the resistors.
- (ii) the total p.d across all resistors in series. = sum of individual p.d. across each resistor.
- (iii) individual p.d. is directly proportional to individual resistors.
- (iv) **equivalent resistance = sum of individual resistances.**

Resistors in parallel

Consider two resistors R_1 and R_2 which are connected in parallel across A and B. A current I enters at A and leaves at B, setting up a potential difference V between them. The current branches into I_1 and I_2 through the two resistors.



$$I = I_1 + I_2$$

$$\frac{\mathcal{E}}{R_{eq}} = \frac{\mathcal{E}}{R_1} + \frac{\mathcal{E}}{R_2}$$

The reciprocal of the equivalent resistance R_{eq} across AB is

$$\frac{1}{R_{eq}} = \frac{1}{R_1} + \frac{1}{R_2}$$

Hence, for n resistors connected in parallel:

$$\boxed{\frac{1}{R_{eq}} = \sum \left(\frac{1}{R_n} \right)}$$

Characteristics of circuits consisting of resistors **in parallel combination** are;

- (i) the potential difference is the same across each of the resistors.
- (ii) total current across parallel combination of resistors = sum of individual currents across each resistor.
- (iii) the larger a resistor the smaller the current through it.
- (iv) equivalent resistance is smaller than the least of the individual resistances.
- (v) **reciprocal of the equivalent resistance = sum of the reciprocals of individual resistances.**



For two particular cases, the formula can be simplified, making for greater speed in calculation:

(a) The combined equivalent resistance of n **EQUAL** resistances R in parallel:

$$R_{eq} = R / n$$

(b) For **two resistors in parallel**:

$$R_{eq} = \frac{R_1 R_2}{R_1 + R_2}$$

Example 5 (N89/II/14)

The circuit diagram shows a network of resistors each of resistance R . What is the effective resistance between the points X and Y?

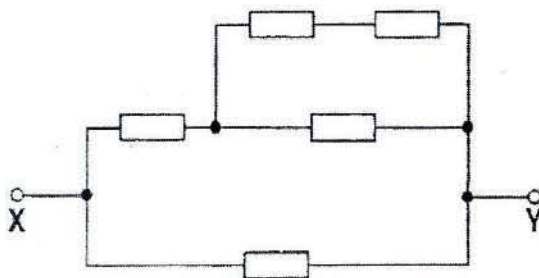
A. $\frac{2R}{7}$

D. $\frac{2R}{3}$

B. $\frac{R}{2}$

E. $\frac{3R}{4}$

C. $\frac{5R}{8}$



$$\frac{1}{2R} + \frac{1}{R} = \frac{1}{R_{eq}}$$

$$R_{eq} = \frac{R_1 R_2}{R_1 + R_2} = \frac{2R}{3}$$

$$\frac{2R}{3} + R = \frac{5R}{3}$$

$$\frac{1}{\frac{5R}{3}} + \frac{1}{R} = \frac{1}{R_{eq}}$$

$$\frac{3}{5R} + \frac{1}{R} = \frac{1}{R_{eq}}$$

$$\frac{8}{5} = \frac{1}{R_{eq}}$$

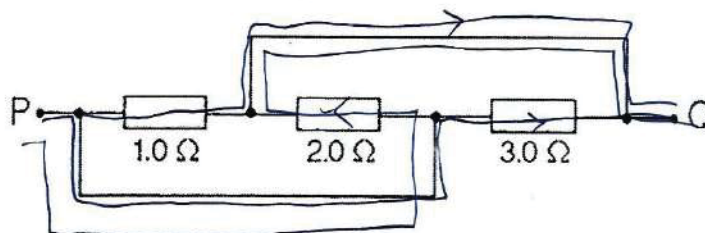
$$R_{eq} = \frac{5R}{8}$$

KIASU
ExamPaper

Example 6 (J89/I/15)

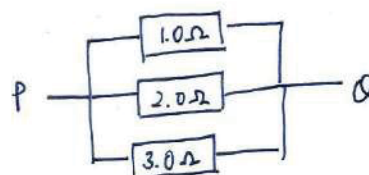
Three resistors are connected as shown in the diagram using connecting wires of negligible resistance. What is the approximate resistance between points P and Q?

- A. 0.5Ω
 B. 0.8Ω
 C. 2.0Ω
 D. 2.2Ω
 E. 3.6Ω



$$\frac{1}{R} = \frac{1}{1} + \frac{1}{2} + \frac{1}{3}$$

$$R = 0.545 \Omega$$



The connecting wires are taken as ideal conductors (i.e., $R = 0 \Omega$), unless otherwise stated. Thus two points joined by a connecting wire would be at the same potential.

13.5.6 Power dissipation in resistors (Joule heating)

From the definitions of resistance and potential difference, we have learnt:

For an electric device of resistance R carrying a steady current I :

The potential difference V across the device is: $V = IR$

The power dissipated by the device is: $P = IV$

Electric components manufactured especially for their resistance are called resistors. These resistors introduced into circuits provide resistances that are large compared with the resistance of the wires and connectors joining together the other circuit components.

When an electric current passes through a resistor, electrical energy is irreversibly transformed into thermal energy (Joule heating). The following two equations describe the rate of transfer of electrical energy to heat generated in a resistor:

$$P = I(IR) = I^2 R$$

This expresses the result first discovered experimentally by Joule:

The rate of production of heat by an electric current in a given resistance is proportional to the square of the current flowing in it.

$$P = \frac{V}{R} V = \frac{V^2}{R}$$

Therefore, we have three expressions for the power production P :

$$P = VI = I^2 R = V^2/R$$



The last two of these formulae ($P = I^2 R = V^2/R$) can apply only to the production of heat in resistors. The formula ($P = VI$) can be applied to determine the rate of production of energy in any form in a circuit component in which a current I flows at a potential difference V .



Students often apply the wrong formula to compute the required power dissipated by a particular component in a circuit as they fail to identify the correct V , I or R in the circuit.

Note that V is the p.d. across the identified component. In questions asking to determine the power loss across the transmission cables, the power loss (through heating) in cables is usually determined using $P = I^2 R$ (where I is the current flowing in the cables and R is the resistance of the cables) as it is often not possible to determine V across the transmission cables.

If the value of V across the transmission cables can be measured, the power loss in the cable can then also be determined using $P = V^2/R$.

Example 7

The power output at a power station is 500 MW at a p.d. of 132 kV. If the total resistance of the complete power line is 4.0Ω , calculate

- the current flowing in the power line,
- the voltage drop caused by the power line, and
- the percentage loss of power in transmission.

Before solving, sketch a schematic representation of this circuit and label all given values in your schematic.

$$(i). \text{ Current, } I = \frac{P}{V} = \frac{500 \times 10^6}{132 \times 10^3} = 3.79 \times 10^3 \text{ A}$$

$$(ii). \text{ Voltage drop across } R_T, V_{\text{cable}} = IR \\ = (3.79 \times 10^3)(4.0) = 15200 \text{ V}$$

$$(iii). \% \text{ loss of power} = \frac{P_{\text{dissipated by cable resistance}}}{P_{\text{supplied by station}}} \times 100\% = \frac{IV_{\text{cable}}}{IV_{\text{supply}}} \times 100\%$$

KIASU
ExamPaper



$$\frac{15200}{132000} \times 100\% = 11.5\%$$

13.6 I-V Characteristics of Various Electrical Components

It is often useful to understand how an electrical component (e.g. resistor, diode, etc) would behave or function in an electrical circuit. This information is often obtained from the I-V characteristics graphs of the required electrical component. Thus it is often useful and critical (for GCE A Level exams) to be able to understand and interpret the I-V characteristics graphs of different electrical components.



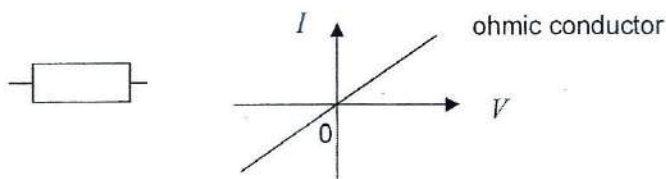
It is important to note that for any I-V characteristics graph, the I represents the current passing through the component and V represents the p.d. across the component. Thus, the resistance of the component (for a particular set of I and V values) is obtained by directly dividing V by I .

An ohmic conductor is one in which the potential difference across it varies proportionally with the current through it.

Metallic Conductor

Metal conductors are ohmic conductors if their physical conditions remain constant. Experimentally, it can be shown that if the temperature of the conductor is kept constant, its resistance will be constant.

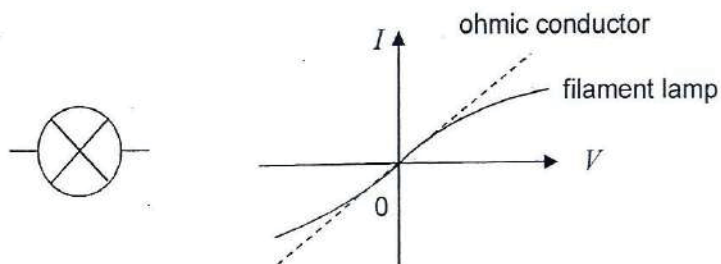
Any increase in the p.d. produces a proportionate increase in the current. The ratio of V to I is constant and the conductor has a single value of resistance. The I-V graph is a straight line passing through the origin.



Microscopically, the resistance of the metal conductor is due to collisions between the free electrons and the metal ions in the lattice. When the temperature of the conductor is kept constant, the amplitude of lattice ion vibrations will remain the same. This means that the hindrance to the flow of electrons is fixed and the resistance will remain constant.

Filament Lamp (case of a metal conductor whose temperature is not controlled)

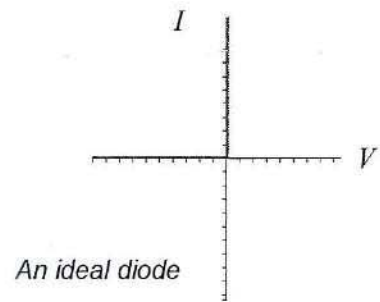
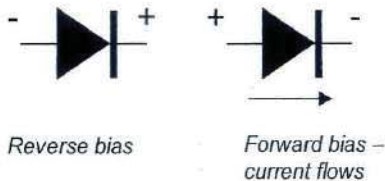
When a larger p.d. is applied, the temperature of the filament increases due to the heating effect of the current. The **resistance of the filament increases with temperature** at larger values of p.d. and current.



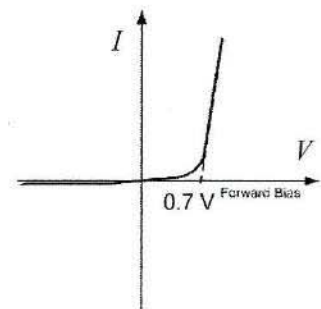
The collisions between the electrons and lattice ions cause energy to be transferred to the lattice ions, enabling the ions to vibrate with larger amplitudes – this implies an increase in the temperature of the filament, and beyond a certain point, the filament glows. The increased vibration of the lattice ions presents a greater hindrance to the flow of electrons, resulting in an increase in the resistance of the filament. Hence at any particular value of p.d. (except at low values of p.d. where there is no significant increase in temperature – any energy gained by the filament is dissipated to the surroundings quickly enough), the current in a filament is lower than that in an ohmic conductor.

Semiconductor Diode

A *p-n junction diode* is a semiconductor which conducts well in one direction of p.d. (forward-bias direction) but conducts poorly when the p.d. is reversed (reverse-bias direction). In an ideal diode (which does not exist in real life), if the diode is reverse biased, the current flowing through it is zero. The ideal diode starts conducting at 0 V and for any positive voltage an infinite current flows and the diode acts like a short circuit (a "perfect" wire).



A real diode needs a minimum positive voltage, called the turn-on voltage, before it conducts any significant current. A silicon diode **requires a p.d. around 0.7 V to conduct**. Once the applied p.d. is **above 0.7 V, the current rises very quickly with the p.d.**, and the diode has very small resistance. If the p.d. is applied in the opposite direction (*reverse biased*), almost zero current flows, and the diode has very high resistance.



Note that if the reverse-biased p.d. is too high, the diode may break down, resulting in a short circuit.

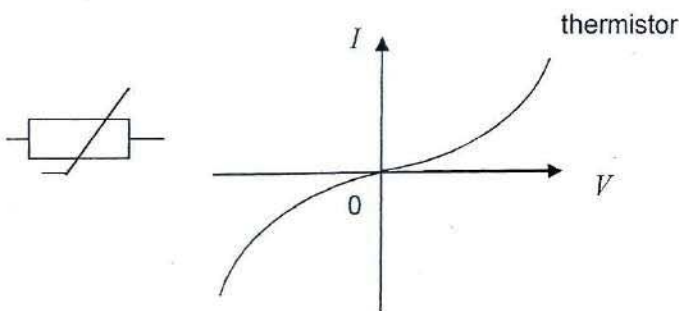
A diode is used as a rectifier for changing an alternating current to a direct current (greater detail is in the topic Alternating Current).

A real diode with turn-on voltage of 0.7 V

(A diode is essentially a p-n junction which will be dealt with in greater detail in the topic Lasers and Semiconductors.)

Thermistor

Usually, a thermistor is a semi-conductor device. Unlike the filament lamp, its **resistance decreases with increasing temperature**.²

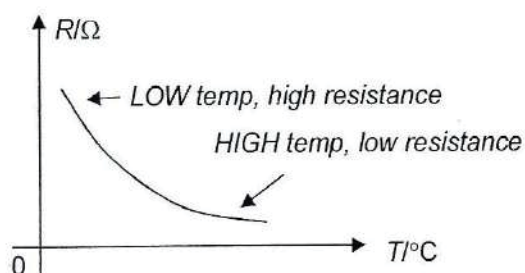


As a larger p.d. is applied, the current through the thermistor increases, causing the thermistor to heat up. Even though there is increased lattice vibrations which increases the hindrance to the flow of the charge carriers (like in the case of the filament), the extra thermal energy supplied to a thermistor *a/so* causes it to release more charge carriers. The outcome is a decrease in the ratio of V to I – a fall in its resistance at higher values of p.d. .

² This is what we call a negative temperature coefficient (NTC) thermistor. There are also positive temperature coefficient thermistors, whose resistance increases with temperature.

The graph below shows the resistance-temperature characteristic of a thermistor.

R - T graph for a thermistor

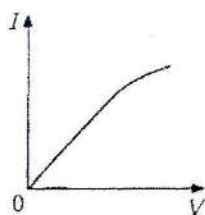


The thermistor's response is rapid and most sensitive to changes to temperature in the region of 200 K to 600K; it is a more limited range compared to thermocouples or resistance thermometers.

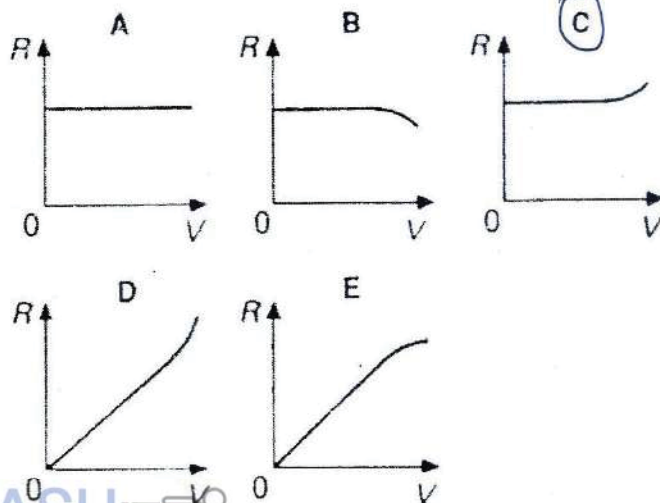
The thermistor is commonly used in very low temperature measurements as the high sensitivity (dR/dT) at low temperature enables more accurate measurements to be made. It can also be used to operate fire alarms and thermostats.

Example 8 (N91/I/16)

The current I flowing through a component vary with the potential difference V across it as shown.



Which graph best represents how the resistance R varies with V ?



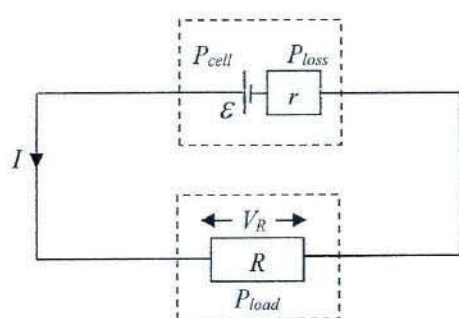
13.7 Internal resistance of a source of e.m.f.

All practical sources of e.m.f. have internal resistance. The chemicals within a cell offer a resistance to current flow, known as the internal resistance of the cell.

13.7.1 Terminal potential difference, e.m.f and internal resistance

Consider a simple circuit with a cell of e.m.f. $\mathcal{E}$ and a load of resistance R . From the energy point of view, the cell is the supplier while the load is the consumer of electrical energy. However, the cell has internal resistance which would consume part of the energy. Hence, less energy is available to drive each coulomb round the external circuit.

A cell can be represented as an ideal source of e.m.f. in series with a resistor representing the internal resistance of the cell. This internal resistance *cannot be* separated out or measured directly.



P_{cell} = power supplied by cell
 P_{loss} = power loss due to internal resistance
 P_{load} = power dissipated in the load

By the law of conservation of energy,

Energy supplied by the cell
 = total energy consumed by the load and internal resistance of the cell

Since energy is supplied continuously, it follows that:

Power supplied by the cell
 = total power consumed by the load and internal resistance of the cell

$$P_{cell} = P_{load} + P_{loss}$$

$$I\mathcal{E} = IV_R + IV_r$$

$$I\mathcal{E} = IV_R + I^2r$$

$$\mathcal{E} = V_R + Ir$$

$$V_R = \mathcal{E} - Ir$$

Thus, the potential difference across the terminals of a cell (terminal p.d.) of a source supplying current to an external circuit is always **smaller than its e.m.f.** due to the presence of the internal resistance, r .

If the circuit is open, no current will flow through the circuit and the terminal p.d. of the source will be equal to the e.m.f. of the source.

The output efficiency of the above circuit is given by:

$$\text{Efficiency} = \frac{\text{useful power output}}{\text{power input}}$$

$$\text{Efficiency} = \frac{P_{\text{load}}}{P_{\text{cell}}} = \frac{I^2 R}{IE} = \frac{IR}{E} = \frac{\left(\frac{E}{R+r}\right)R}{E} = \frac{R}{R+r}$$

Example 9

A complete circuit consists of a 24.0 V battery, a 5.60 Ω resistor and a switch. The internal resistance of the battery is 0.28 Ω . The switch is opened. What does a voltmeter read when placed

- (a) across the terminals of the battery?
- (b) across the resistor?
- (c) across the switch?

Repeat (a), (b) and (c) for the case when the switch is closed.

Open,

- a) There is no current, so no p.d. across r: voltmeter reading is 24.0 V
- b) There is no current so no p.d. over R: voltmeter reading is 0V.
- c) Potential at poles of battery will be the same as that on each side of the switch: voltmeter reading is 24.0V.

a) $E = IR + Ir$

$$24.0 = I(R + r)$$

$$I = \frac{24.0}{(5.60 + 0.28)}$$

$$= 4.082 \text{ A}$$

$$V = E - Ir = 24.0 - (4.082 \times 0.28) = 22.9 \text{ V}$$

- b) Potential drop over R, V_R = potential drop across the terminals of the battery: voltmeter reads 22.9 V

- c) The switch is closed, so there is no potential drop across the switch: voltmeter reads 0V.

13.8 The Potential Divider Principle

A potential divider is an arrangement of resistors which is used to supply varying levels of p.d. to a circuit from fixed e.m.f. sources.

Consider two resistors connected in series.

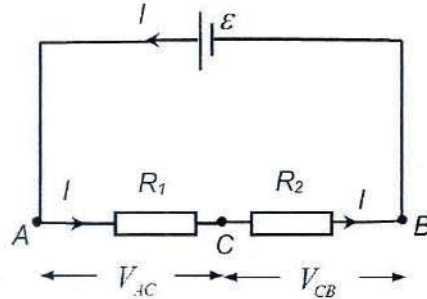


Figure 13.8.1

The current I flowing in the circuit is given by

$$I = \frac{\mathcal{E}}{R_1 + R_2}$$

Therefore, the potential difference across R_1 , V_{AC} is

$$\begin{aligned} V_{AC} &= IR_1 \\ &= \left(\frac{\mathcal{E}}{R_1 + R_2} \right) R_1 \\ &= \frac{R_1}{R_1 + R_2} \mathcal{E} \end{aligned}$$

The potential difference across R_2 , V_{CB} is

$$\begin{aligned} V_{CB} &= IR_2 \\ &= \left(\frac{\mathcal{E}}{R_1 + R_2} \right) R_2 \\ &= \frac{R_2}{R_1 + R_2} \mathcal{E} \end{aligned}$$

Hence, by adjusting the ratio of R_1 to R_2 , we can obtain the desired potential difference across AC.



Note: The supply voltage is divided proportionately between the two resistors if the same current flows in both resistors. The fraction depends on the **RATIO of the resistors** and not the absolute values of the resistances. It is assumed that this cell has no internal resistance. If the cell has internal resistance, a proportionate amount of voltage will drop across the internal resistance.



Instead of two resistors, a **rheostat** can be used as in Figure 13.8.2 as a source of variable potential difference.

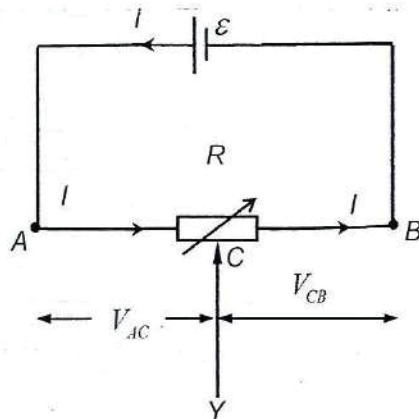


Figure 13.8.2

By sliding the contact/slider Y on the rheostat R , V_{AC} can be made to vary from 0 (when Y is tapped at the point of zero resistance) to ε (when Y is tapped at the point of maximum resistance), assuming the driver cell has negligible internal resistance.

13.8.1 Use of thermistors and light dependent resistors in potential dividers

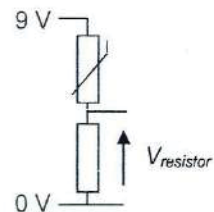
Thermistors

One of the resistors in Figure 13.8.1 can be substituted with a thermistor. When the temperature increases, the resistance of the thermistor decreases, which in turn causes a decrease in the p.d. across the thermistor

($V_{\text{thermistor}} = \frac{R_{\text{thermistor}}}{R_{\text{thermistor}} + R_{\text{resistor}}} \varepsilon$) and an increase in the p.d. across the other

resistor

($V_{\text{resistor}} = \frac{R_{\text{resistor}}}{R_{\text{thermistor}} + R_{\text{resistor}}} \varepsilon$).



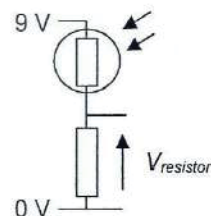
One application of a thermistor is its use as a temperature sensor in a fire alarm. When hot conditions are detected, V_{resistor} increases, causing a fire alarm circuit to be triggered.

Light Dependent Resistors

One of the resistors in Figure 13.8.1 can be substituted with a light dependent resistor (LDR). A LDR is a device whose resistance decreases as the intensity of light incident on it increases. Under bright conditions, the p.d. across the LDR

($V_{\text{LDR}} = \frac{R_{\text{LDR}}}{R_{\text{LDR}} + R_{\text{resistor}}} \varepsilon$) decreases, while the p.d. across the other resistor in the

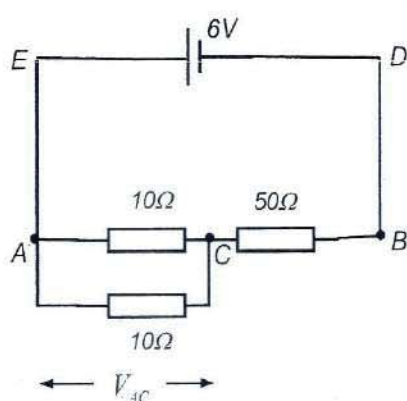
circuit increases ($V_{\text{resistor}} = \frac{R_{\text{resistor}}}{R_{\text{LDR}} + R_{\text{resistor}}} \varepsilon$).



One could use a LDR to control lighting circuits which are switched on automatically in the evening when illumination falls below a certain intensity.

Example 10

(a) Determine the potential difference across the junction AC in the following circuit.



$$a) V_{AC} = \frac{\left(\frac{1}{10} + \frac{1}{10}\right)^{-1}}{\left(\frac{1}{10} + \frac{1}{10}\right)^{-1} + 50} (6) = 0.55V$$

$$b) V_{AC} = \frac{\left(\frac{1}{10} + \frac{1}{10} + \frac{1}{200}\right)^{-1}}{\left(\frac{1}{10} + \frac{1}{10} + \frac{1}{200}\right)^{-1} + 50} (6) = 0.53V$$

$$c) I = \frac{E}{R_{total}} = \frac{6}{\left(\frac{1}{10} + \frac{1}{10}\right)^{-1} + 50 + 10}$$

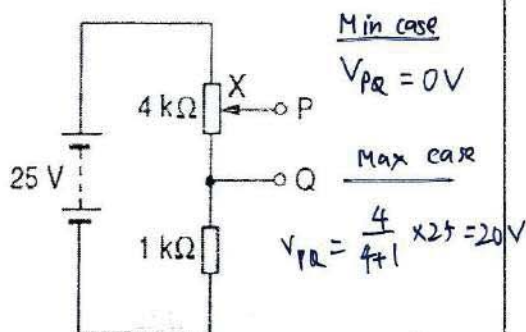
(b) Determine the reading of a voltmeter of resistance 200 Ω placed across AC. $= 0.092A$

(c) Determine the reading of an ammeter of resistance 10 Ω placed in the path BD.

Example 11 (J86/I/15)

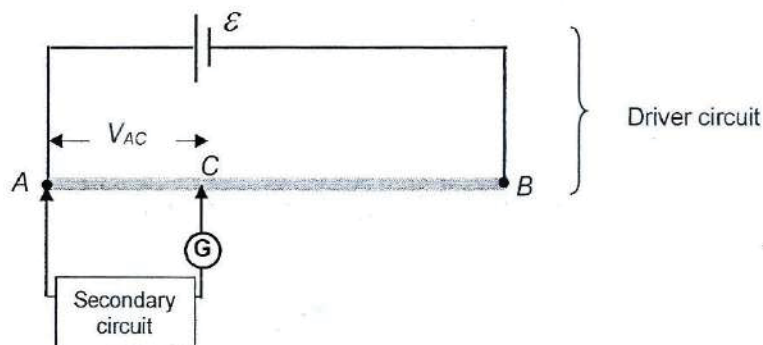
The diagram on the right shows a potential divider circuit which, by adjusting the position of the contact X, can be used to provide a variable potential difference between the terminals P and Q. What are the limits of this potential difference?

- A 0 and 25 mV
- B 5 mV and 25 mV
- C 0 and 20 V**
- D 0 and 25 V
- E 5 V and 25 V



13.9 The Potentiometer

The potentiometer comprises a resistance wire AB and a jockey that is connected to the wire which can be shifted up and down the wire. It is usually connected to a driver circuit and a secondary circuit (the secondary circuit will be elaborated in a later section). In its most basic form, the driver circuit consists of a driver cell, $\mathcal{E}$ and a long resistance wire, shown as AB below.



A voltmeter connected across AC will measure a fraction of the potential difference $\mathcal{E}$. The fraction of $\mathcal{E}$ measured out depends on where point C – a metallic contact called a jockey – is placed along AB.

The principle on which a potentiometer works is exactly the same as for a potential divider. Assuming that the resistance wire AB has a uniform cross-sectional area and a length l_{AB} , the resistance R_{AC} will be proportional to the length l_{AC} , since $R = \frac{\rho l}{A}$, or $R \propto l$. Therefore,

$$V_{AC} = \frac{R_{AC}}{R_{AC} + R_{CB}} \mathcal{E} = \frac{l_{AC}}{l_{AB}} \mathcal{E}$$

A Balanced Potentiometer

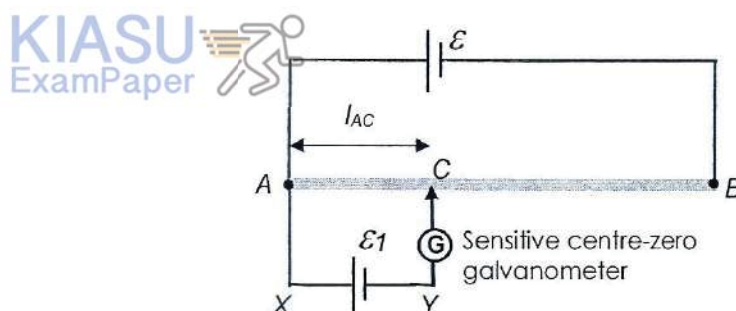
The potentiometer is said to be **balanced** when the jockey is at a position on AB such that there is **NO current** through the galvanometer (essentially an ammeter). This means that there is no current or charge flow between the driver circuit and the secondary circuit, i.e. the driver circuit and secondary circuit are electrically isolated. The applications of the potentiometer that we will discuss in the next section rest on finding this balanced point.

13.9.1 Applications of the potentiometer

The secondary circuit can take on many forms, depending on the application of the potentiometer.

Application 1: To determine the unknown e.m.f. of a cell

The secondary circuit is just composed of the unknown e.m.f. source $\mathcal{E}_1$.



By adjusting the length l_{AC} , we change the p.d. V_{AC} .

At the point where the galvanometer shows null deflection (no current flows through it and the potentiometer is balanced),

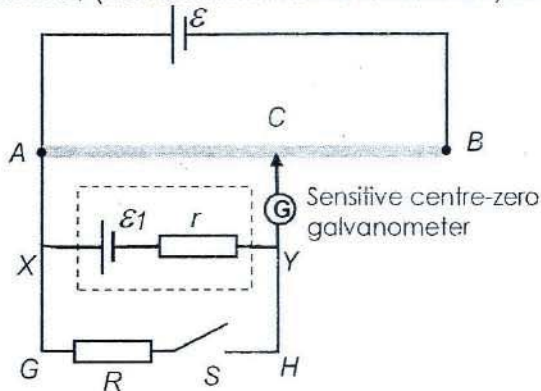
$$\begin{aligned}\mathcal{E}_1 &= V_{XY} = V_{AC} \\ \mathcal{E}_1 &= \frac{l_{AC}}{l_{AB}} \mathcal{E}\end{aligned}$$

The length l_{AC} is known as the balanced length. By measuring l_{AC} , with known l_{AB} and $\mathcal{E}$, $\mathcal{E}_1$ can be found.

Since there is no current flowing through the galvanometer, the potentiometer essentially functions as an ideal voltmeter!

Application 2: To determine the unknown internal resistance of a cell

A driver cell with e.m.f. $\mathcal{E}$ has negligible internal resistance. The secondary circuit consists of an unknown e.m.f. $\mathcal{E}_1$ (with unknown internal resistance r) connected in series with an external resistor R .



When **switch S is open**, there is no current flowing in XGHY.

When the galvanometer shows null deflection (no current flowing through AXYC),

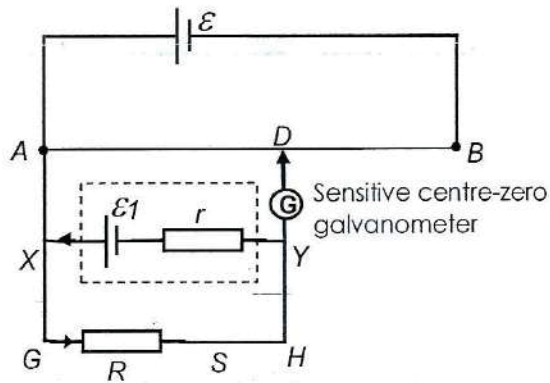
$$V_{XY} = \mathcal{E}_1 \text{ (since there is no p.d. across } r\text{)}$$

Since $V_{AC} = V_{XY}$, $V_{AC} = \mathcal{E}_1$

From the driver circuit, $V_{AC} = \frac{l_{AC}}{l_{AB}} \mathcal{E}$

Therefore, $\mathcal{E}_1 = \frac{l_{AC}}{l_{AB}} \mathcal{E}$ [1]

When we close the switch S, a current flows in the circuit XGHY.



At a new balanced length AD (no current flows **between** the driver and secondary circuits, even though a current flows in the driver circuit and a (possibly different) current flows in the secondary circuit),

$$V_{XY} = V_{GH} = \frac{R}{R+r} \varepsilon_1$$

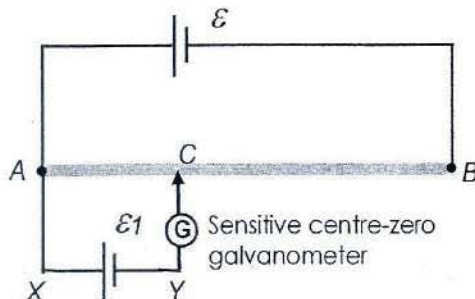
Since $V_{AD} = V_{XY}$,

$$\frac{l_{AD}}{l_{AB}} \varepsilon = \frac{R}{R+r} \varepsilon_1 \quad [2]$$

Since ε , R , l_{AC} , l_{AD} and l_{AB} are known, ε_1 and r can be calculated using equations [1] and [2].

Example 12

In the circuit shown below, cell $\mathcal{E}$ has an e.m.f. of 2.0 V and negligible internal resistance. AB is a uniform wire of length 1.00 m and resistance 5.0 Ω . Cell $\mathcal{E}_1$ has an e.m.f. of 1.5 V and internal resistance of 0.80 Ω . Calculate the length AC required to produce zero deflection in the galvanometer

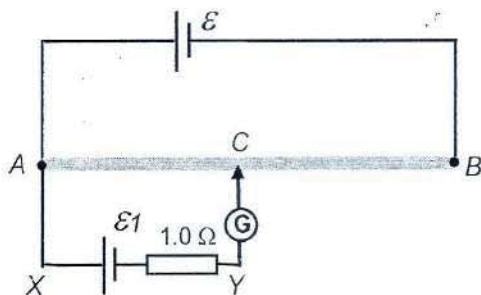


(i) in the circuit as shown above,

$$V_{XY} = \mathcal{E}_1 \quad \mathcal{E}_1 = \frac{l_{AC}}{L} \mathcal{E}$$

$$l_{AC} = \frac{\mathcal{E}_1 L}{\mathcal{E}} = \frac{1.5}{2.0} \times 1.00 = 0.75 \text{ m}$$

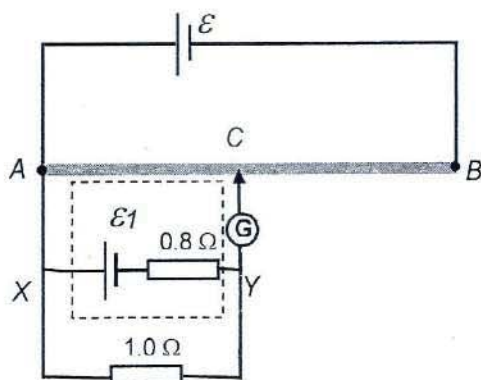
(ii) when a resistor of 1.0 Ω is placed in series with $\mathcal{E}_1$ as shown below.



Meter has zero deflection, no current flows thru AXYC.

$$l_{AC} = 0.75 \text{ m}$$

(iii) when a resistor of 1.0 Ω is connected with $\mathcal{E}_1$ as shown below.



At balanced pt,
No current flows AX & YC

$$V_{AC} = V_{XY}$$

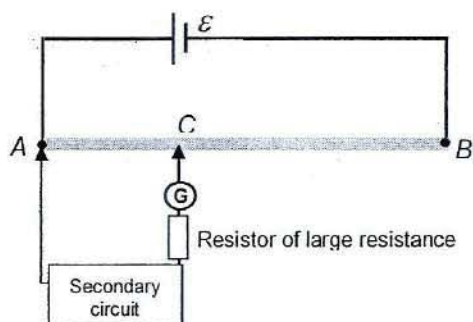
$$V_{XY} = \frac{1.0}{1.0 + 0.8} \times 1.5 = 0.833 \text{ V}$$

$$V_{XY} = V_{AC} = \frac{l_{AC}}{l_{AB}} \mathcal{E}$$

$$l_{AC} = \frac{0.833}{2.0} \times 1.00 = 0.42 \text{ m}$$

Practical pointers on the potentiometer

- The e.m.f.s of both the driver cell and the unknown cell should be constant throughout. If not, the balance length will fluctuate.
- The jockey should not be scraped along wire AB to determine the balance point, because doing so could damage the uniformity of the wire. Instead the jockey gently taps along the wire AB.
- A very sensitive galvanometer is needed if precise locations of the balance points are to be made. Such a galvanometer should be protected, by a large (typically $1\text{ M}\Omega$) series resistance, from the relatively high current which would flow through it when the jockey is far away from the balanced point. The presence of this resistor will not affect the balance length since at balanced point, no current flows through the galvanometer (and the resistor of large resistance).

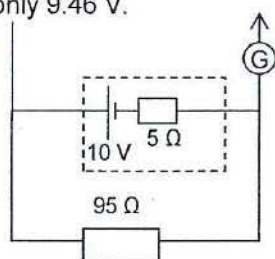


Advantages of using a potentiometer to measure voltage

- It draws no current. Therefore, it functions as an ideal voltmeter.

When used to measure an unknown e.m.f., the value obtained at balanced point is the true e.m.f. If a practical voltmeter of resistance $1000\ \Omega$ is used across an e.m.f. source of 10 V and internal resistance $5\ \Omega$, the practical voltmeter will read only $V = \frac{1000}{1005} \times 10 = 9.95\text{ V}$.

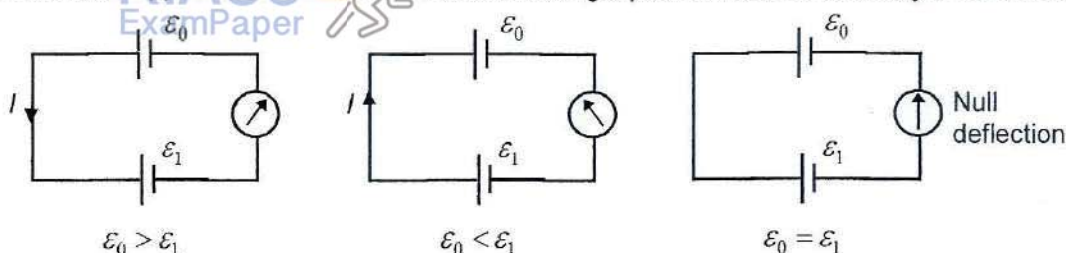
- When used to measure the terminal p.d., the value obtained at balanced point is the true terminal p.d. – which would be 9.5 V for the circuit below. If a practical voltmeter of resistance $1000\ \Omega$ is used instead to measure the terminal p.d., the effective external resistance would be $(\frac{1}{95} + \frac{1}{1000})^{-1}$ and the voltmeter would read only 9.46 V .



Disadvantages of using a potentiometer to measure voltage

- It is slow in operation. It cannot be used to monitor rapidly changing voltages.
- The resistance wire may not be uniform due to wear and tear.

It should be appreciated that the measurement using a potentiometer is basically a method of comparison.



APPENDIX A – Two different arrangements of voltmeter and ammeter to find a more accurate value of the resistance of an unknown resistor.

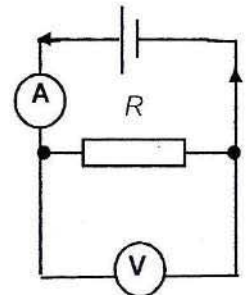
If the meters are not ideal, the significance of the error in the meter readings can be reduced by using the following circuit arrangements:

If the unknown R is expected to be low

The voltmeter reads the correct value of p.d. across R .

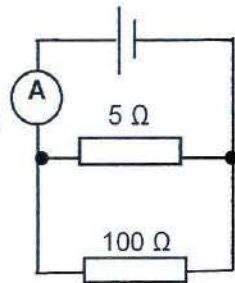
While the ammeter reads the sum of the current flowing through both R and the voltmeter, because R is low, most of the current actually passes through R rather than the voltmeter – hence the error in the ammeter reading will not be too great.

If this circuit were used for a high R (perhaps comparable to the resistance of the voltmeter) instead, the ammeter will not be measuring a value of current close to the current actually flowing through R – the error in the ammeter reading will be more significant (the voltmeter still reads the p.d. across R).

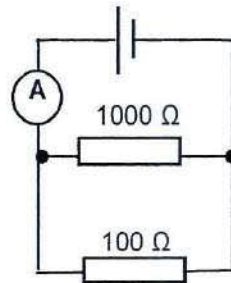


No current passing through ideal voltmeter

Ammeter reading deviates from the true value by about 5%



Voltmeter with resistance



Voltmeter with resistance

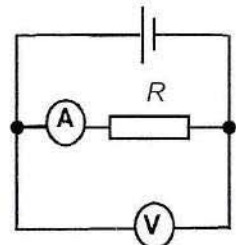
Ammeter reading deviates from the true value by about 91%

If the unknown R is expected to be high

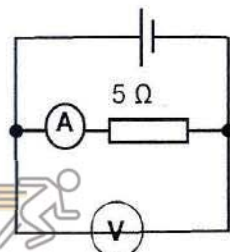
The ammeter reads the correct value of current flowing through R .

While the voltmeter reads the sum of the p.d. across both R and the ammeter, because R is high, most of the p.d. actually falls across R – hence the error in the voltmeter reading will not be great.

If this circuit were used for a low R (perhaps comparable to the resistance of the ammeter) instead, the error in the voltmeter reading will be more significant (the ammeter still reads the current through R).

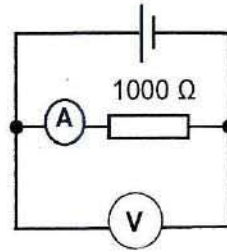


Ammeter with 1 Ω resistance.



Voltage reading of 5 Ω resistor would deviate by 20% from the true value.

Ammeter with 1 Ω resistance



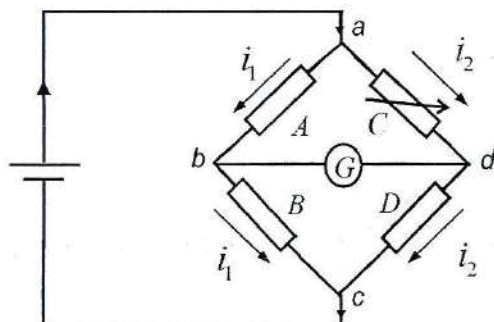
Voltage reading of 1000 Ω resistor would deviate only by 0.5% from the true value.

KIASU
ExamPaper

APPENDIX B – The Wheatstone bridge (Good to know)

Unknown resistances can be accurately measured using a circuit known as a Wheatstone bridge.

This circuit consists of the unknown resistance A, three known resistors B, C and D (where C is a variable resistor), a galvanometer, and a source of e.m.f.



Since point *a* is common and the potential at *b* = the potential at *d*, the potential difference across *ab* = potential difference across *ad*. Likewise, the potential difference across *bc* = potential difference across *dc*.

Resistor C is varied until the galvanometer reading is zero (null deflection), i.e. there is no current flowing through the galvanometer between points *b* and *d*. Under this condition of null deflection, the bridge is said to be balanced, and the current which flows in A flows in B as well; the current which flows in C flows in D as well.

$$i_1 A = i_2 C \quad (1)$$

$$i_1 B = i_2 D \quad (2)$$

$$(1)/(2) \quad \frac{A}{B} = \frac{C}{D}$$

Since B, C and D are known, A can be found.

Advantages

- Variation in the e.m.f. of the supply does not affect the balance point.

Disadvantages

- This method is unreliable for finding small resistances ($< 1 \Omega$) because the unknown resistances of the (real) wires used to connect the resistors into the bridge are significant.
- This method is insensitive for high resistances ($\sim 10^6 \Omega$) unless a highly sensitive galvanometer is used. This is because the presence of high resistances in the circuit causes the current through the galvanometer to be near zero even when the potentials of points *b* and *d* are far from each other.

Topic 14: Electromagnetism

C2 2017

H2 Physics Syllabus 9749

Content

- Concept of a magnetic field
- Magnetic fields due to currents
- Force on a current-carrying conductor
- Force between current-carrying conductors
- Force on a moving charge

Learning Outcomes

Candidates should be able to:

- (a) show an understanding that a magnetic field is an example of a field of force produced either by current-carrying conductors or by permanent magnets
- (b) sketch flux patterns due to currents in a long straight wire, a flat circular coil and a long solenoid
- (c) use $B = \mu_0 I / 2\pi d$, $B = \mu_0 NI / 2r$ and $B = \mu_0 nI$ for the flux densities of the fields due to currents in a long straight wire, a flat circular coil and a long solenoid respectively
- (d) show an understanding that the magnetic field due to a solenoid may be influenced by the presence of a ferrous core
- (e) show an understanding that a current-carrying conductor placed in a magnetic field might experience a force
- (f) recall and solve problems using the equation $F = BIl \sin\theta$, with directions as interpreted by Fleming's left-hand rule
- (g) define magnetic flux density and the tesla
- (h) show an understanding of how the force on a current-carrying conductor can be used to measure the flux density of a magnetic field using a current balance
- (i) explain the forces between current-carrying conductors and predict the direction of the forces
- (j) predict the direction of the force on a charge moving in a magnetic field
- (k) recall and solve problems using the equation $F = BQv \sin\theta$
- (l) describe and analyse deflections of beams of charged particles by uniform electric and uniform magnetic fields
- (m) explain how electric and magnetic fields can be used in velocity selection for charged particles.

List of symbols

The following are the main symbols used in diagrams:

- ⊗ Into the page
- ⊙ Out of the page



14 Introduction

More than 2000 years ago, the Chinese discovered magnetic phenomena and have since been using them for navigation. However, it was only in the 1700s that magnetic properties and phenomena were studied in greater depth. Major discoveries made in the last 300 years have made a significant impact on our lives today. Scientists who have contributed significantly to our understanding of magnetic phenomena include Ampere, Oersted, Gauss, Henry and Faraday.

Magnetism has many useful applications in our everyday lives. Earth's magnetic field not only allows us to find our way around, but also enables birds to navigate when they migrate South during winter in the North. Nuclear Magnetic Resonance (NMR) which makes use of atoms' magnetic moments allows us to capture images of our body. Mobile phones, speakers and microphones make use of the concept of electromagnetic induction. The Great Industrial Revolution would not have been possible without knowing that the cutting of magnetic flux can generate electricity.

In the lectures on Electric Field, we have studied forces due to electric charges. You have learnt that there are two types of charges: positive and negative. In this topic, you will learn about forces due to magnetic fields, how magnetic fields can be generated and subsequently how electricity can be generated using magnetic fields. We will again be dealing with charges, as moving charges generate a magnetic field.

14.1 The Magnetic Field and Magnetic Flux Density

Just like the gravitational and electric fields, the magnetic field is an example of a '*field of force*'. The concept of a field illustrates '*action at a distance*', which means that a suitable object placed in the field can experience a force. A region of space is said to be covered with a magnetic field when a magnetic material or a moving charge/current-carrying conductor placed in that region of space experiences a magnetic force.

As with electric fields, we can illustrate magnetic fields by drawing magnetic field lines. An easy way to draw magnetic field lines is to scatter fine iron filings around a magnet. Each iron filing is magnetized and acts like a tiny compass needle. Thus, the filings rearrange themselves along the magnetic field lines.

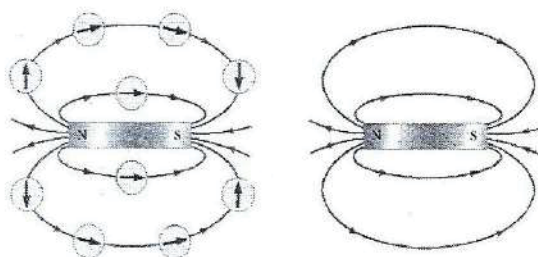
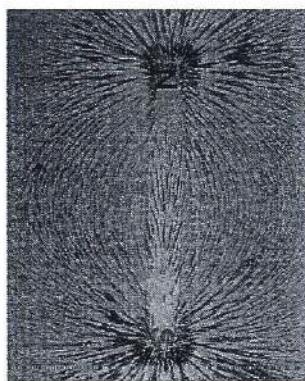


Figure 14.1.1 Plotting of magnetic field lines of a magnet using a compass.

Note that,

- Magnetic field lines 'appear' to originate from the north pole and end on the south pole of a magnet. But to be exact, **magnetic field lines have no start or end point**. The field lines actually continue on within the magnet to form a complete loop.
- The *direction of a magnetic field at any point* is indicated by the **tangent to the magnetic field line** at that point. The direction of the magnetic field at any point in space is the direction indicated by the north pole of a small compass needle placed at that point.
- Magnetic field lines **do not cross each other**.

- The **strength of the magnetic field** is indicated by the **density of the field lines**; the magnetic field is stronger in regions where the field lines are relatively close together and weaker where they are relatively far apart.

Are magnetic field lines of any use?

For now, we use the **direction of the field lines at a point to define the direction of a vector quantity called magnetic flux density B** . As we shall soon discover, this quantity will come in very useful when we try to form an equation to calculate the force that the field can exert.

Further, the **magnitude of this quantity will indicate the strength of the magnetic field**. Nearer to the source, the magnetic field is stronger and thus B will have greater magnitude. The SI unit for magnetic flux density is the **tesla (T)**.

14.2 Currents can generate Magnetic Fields

Before the 1800s, magnetism and electricity were considered to be two independent topics. Although there were a number of scientists who thought the two topics were related, they were unsuccessful in finding the link. It was only in 1820 that a Danish physicist, Hans Christian Oersted, discovered the link. He observed that a compass needle swung almost perpendicular to a current-carrying conductor as if attracted by a powerful magnet. This discovery prompted Ampère to work out the mathematical relationship between currents and magnetic fields known as Ampère's force law (not required in the 'A' level syllabus). It also inspired Faraday's work on electromagnetic induction. Oersted's discovery, which linked the motion of electric charges with the creation of a magnetic field, marked the beginning of an important discipline called electromagnetism.

Two important observations were made about the **magnetic field due to current-carrying conductors**:

- The direction of the magnetic field is always perpendicular to the direction of the current.
- The strength of the magnetic field produced is dependent on the magnitude of the current.

Previously, we detect the presence of a magnetic field of a magnet by observing how the iron filings align themselves when sprinkled around the magnet. A similar effect is observed when we sprinkle iron filings around a current carrying wire!

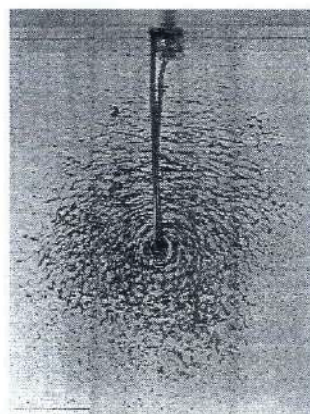
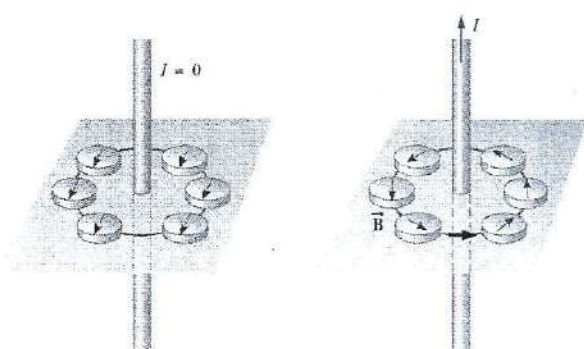


Figure 14.2.1 Compasses deflected when the current is on.

The direction of the magnetic field lines can be easily remembered using the **Right-Hand Grip Rule**.

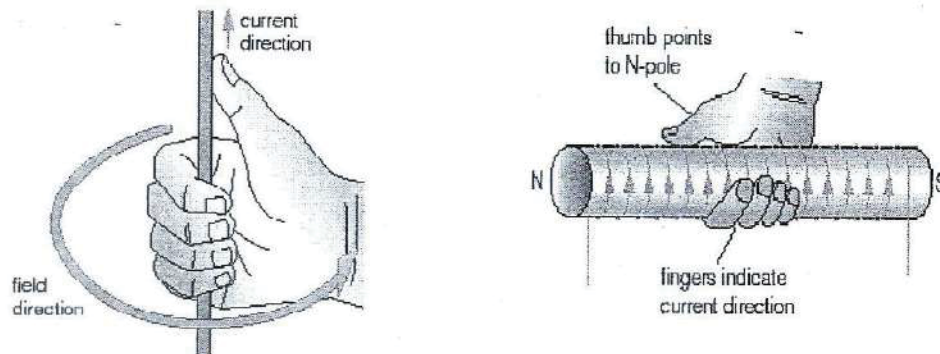
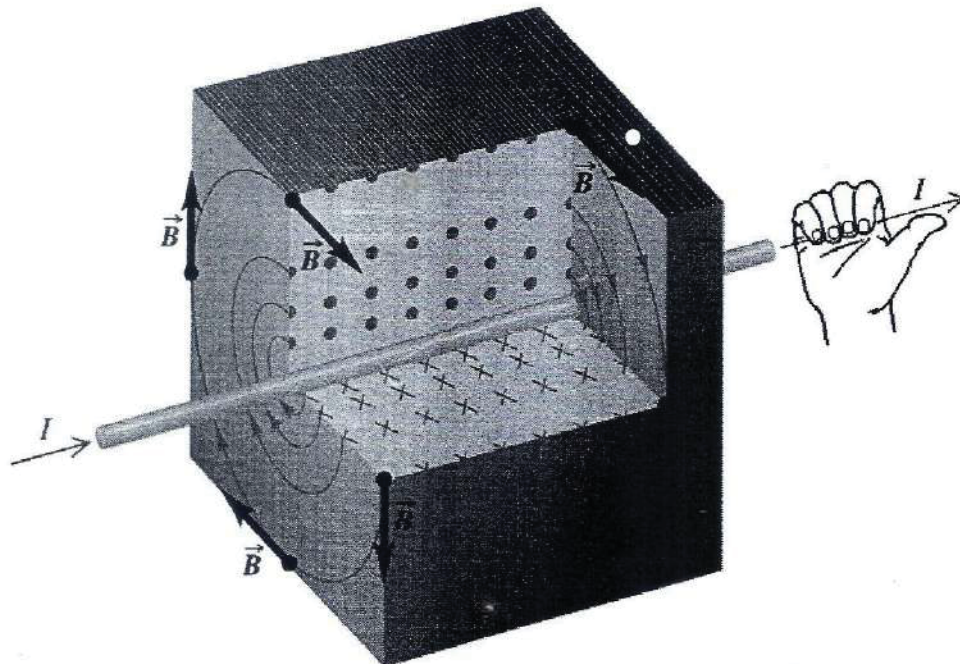


Figure 14.2.2 Using the right-hand grip rule to determine the direction of the magnetic field due to a current-carrying conductor.

14.2.1 Magnetic-Field Patterns

14.2.1.1 Long Straight Current-Carrying Conductor

Figure 14.2.3 shows the magnetic-field pattern produced by a long straight current-carrying conductor.

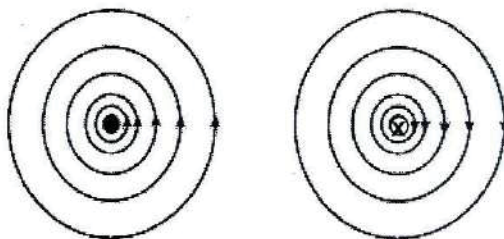


Copyright © Addison Wesley Longman, Inc.

Figure 14.2.3 Magnetic-field pattern due to a long straight current-carrying conductor.

Helpful
Tips

1. The magnetic field lines form concentric circles around the straight wire. Given a choice, always draw the plan view as shown below.



A circle with a DOT shows that the current is coming OUT from the plane

A circle with a CROSS shows that the current is moving IN TO the plane

2. When drawing the lines, it is extremely important to show the field lines spreading further and further apart as r increases. This is because the B-field is getting weaker as we move away from the source (i.e. the wire).

The magnetic flux density, B , at a distance, r , from a long straight wire carrying a current, I , is given by

$$B = \frac{\mu_0 I}{2\pi r}$$

where μ_0 is the **permeability of free space** which has the value $4\pi \times 10^{-7} \text{ H m}^{-1}$.

Thus B is **directly proportional** to I but is **inversely proportional** to the perpendicular distance r from the wire.



14.2.1.2 Flat Circular Coil

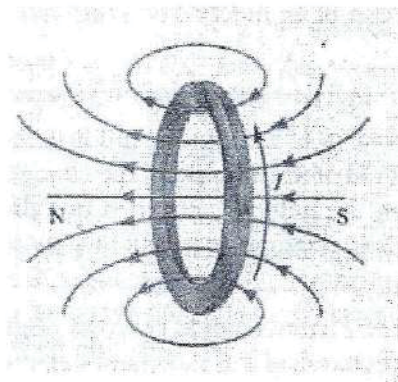
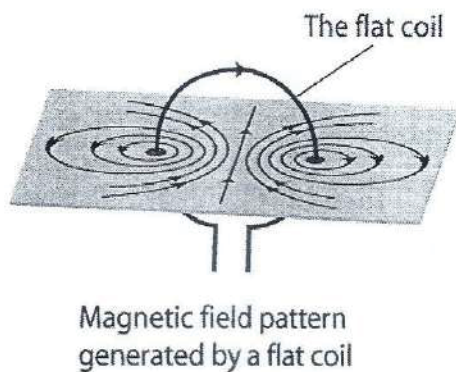
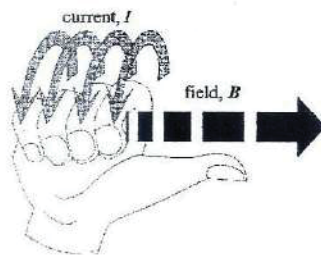


Figure 14.2.4 Magnetic-field pattern of current flowing in circular loop.

The right-hand grip rule can still be used to determine the direction of the magnetic field lines produced by a current in circular coils. In general, there are 2 different methods that can be used:

1. For the straight-current scenario, the thumb represents the current and the coiling fingers representing the B-field.
2. For "circular" currents, we can exchange the fingers and the thumb to let the coiling fingers represent the direction of the currents in the circular coils and the thumb the direction of the magnetic-field lines passing through the cross-sectional area of the circular coils



1. Using the thumb to represent the B-field also allows us to determine the "north" pole of the circular coil, since B-field lines originate from the "north" pole of the magnet. This will be very useful later.
2. It is sometimes difficult to tell from the diagram which part represent the "front" of the coils. Standard conventions include using a hard line for the "front" and a dotted line for the "back" and using a thicker/bolder line for the "front" of the coil.

The magnetic flux density, B , at the centre of a flat circular coil of N turns, radius, r , and carrying current, I , is given by

$$B = \frac{\mu_0 NI}{2r}$$

where μ_0 is the permeability of free space which has the value $4\pi \times 10^{-7} \text{ H m}^{-1}$.

Again, we observe the same relationship as the magnetic field due to a long straight current-carrying conductor, that is, B is directly proportional to I but is inversely proportional to the radius of the circular coil, r

14.2.1.3 Long Solenoid

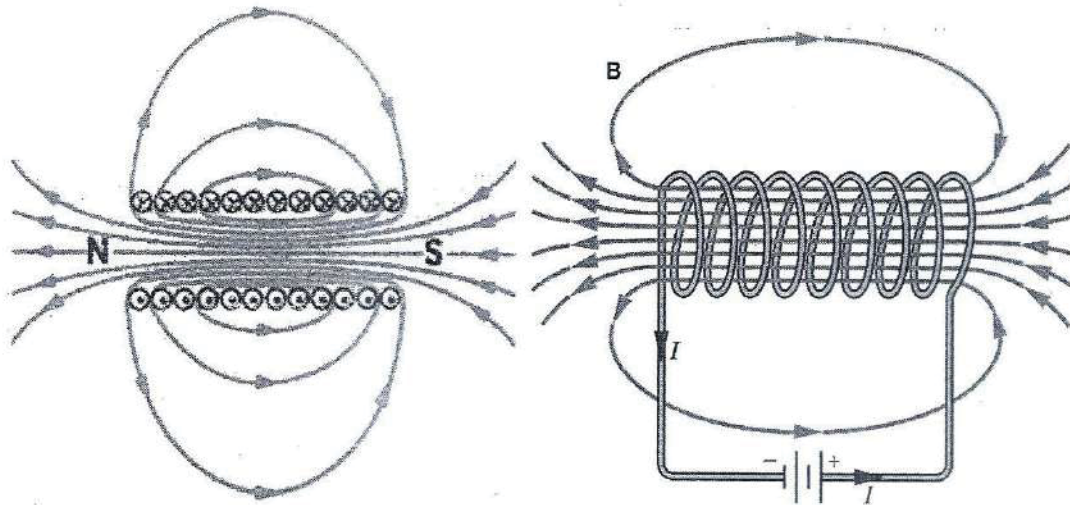


Figure 14.2.5 Magnetic-field pattern of a current-carrying solenoid.



1. The magnetic field produced by a long solenoid is **uniform** for the length of the solenoid except for the ends of the solenoid. This is represented by the **parallel, equally-spaced lines along the axis of the solenoid**.
2. The solenoid also has a "north" and a "south" pole. In fact, its field pattern is very similar to that of a short bar magnet.
3. The magnetic flux density **at the ends is half the magnetic flux density at the center** of the solenoid.

The magnetic flux density, B , at the **centre of a long solenoid** with n turns per unit length and carrying current, I , is given by

$$B = \mu_0 n I$$

where μ_0 is the **permeability of free space** which has the value $4\pi \times 10^{-7} \text{ H m}^{-1}$.

Here, B is **directly proportional to I and n** (the number of turns per unit length). Note that n (lower-case n) is the *number of turns per unit length*, whereas N (capital N) is the *total number of turns*.

One way of strengthening the magnetic field of a solenoid is to insert a piece of soft iron (ferrous core) into the solenoid. The magnetic flux density at the poles of the solenoid will be modified.

$$B = \frac{1}{2} \mu_0 n I \Rightarrow B = \frac{1}{2} \mu n I$$

In this case, the soft iron, which can be permanently magnetized, increased μ_0 to μ . The magnetic domains in the soft iron causes the magnetic field to strengthen. Inserting the soft iron causes the permeability of free space, μ_0 , to change to permeability of iron, μ . Materials that add to magnetic field in this manner are described as **ferromagnetic**.

Examples of ferromagnetic materials include soft iron, steel, cobalt and nickel. They could all be magnetized.

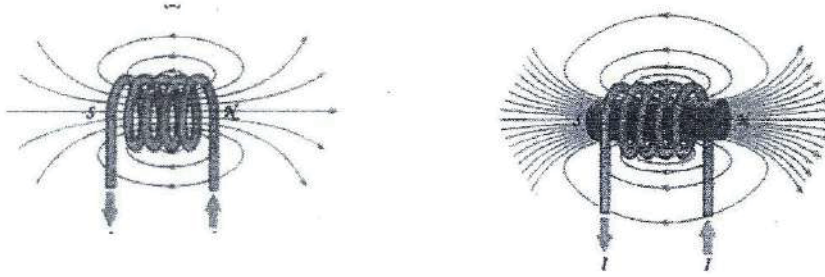
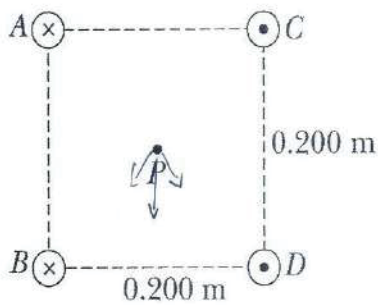


Figure 14.2.6 Increasing field lines due to ferrous core

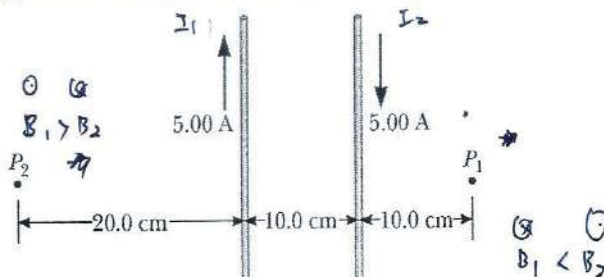
Worked Example 1

Determine the direction of the magnetic field at the center of the square assuming all currents are of the same value.



Worked Example 2

(a) Determine the direction of the magnetic field at points P_1 and P_2 .



At P_1

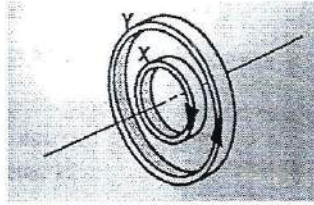
$$B_{\text{net}} = B_2 - B_1$$

$$= \frac{\mu_0}{2\pi} \left(\frac{I_2}{r_2} - \frac{I_1}{r_1} \right)$$

(b) Calculate the magnitude of the resultant magnetic flux density at P_1 .

Worked Example 3 (J93/1/8)

Two coils X and Y, each with 100 turns, are arranged as shown in the diagram.



X has a radius 0.050 m and carries a current of 3.0 A. Y has radius 0.10 m and carries a current of 6.0 A in the opposite direction to X.

What is the magnitude of the total magnetic flux density at the centre of the coils?

$$B_x = \frac{\mu_0 N I_x}{2\pi r_x} \quad B_y = \frac{\mu_0 N I_y}{2\pi r_y}$$

Taking to the right as +ve, total magnetic flux density at centre of coil

$$\begin{aligned} B_{\text{net}} &= B_x - B_y \\ &= \frac{\mu_0 N}{2\pi} \left(\frac{3.0}{0.050} - \frac{6.0}{0.10} \right) = 0 \text{ T} \end{aligned}$$

14.3 Currents Experiencing a Force when in a Magnetic Field

Since currents can generate a magnetic field and magnetized conductors behave like magnets, maybe a current-carrying wire will experience a magnetic force when placed beside a magnet too. The figure below shows the results of such an experiment.

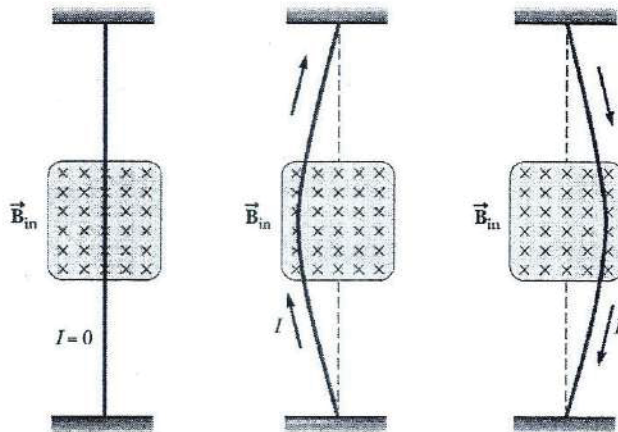


Figure 14.3.1 Each sketch shows the direction of the force exerted on a current-carrying wire placed near a magnet.

Observations from this experiment show that there is a relationship between the magnetic flux density of the magnetic field present and the force experienced by the current-carrying wire.

Maximum force is experienced by the wire when the magnetic flux density B and the wire are **perpendicular to each other**. The direction of the force that results, is perpendicular to both B and the wire. Quantitatively, in this orientation, we have

$$F_B = BIL$$

where F_B is the force exerted on the wire, I is the current flowing in the wire and L is the length of the wire.

The relative directions are easily remembered by **Fleming's Left-Hand Rule**,

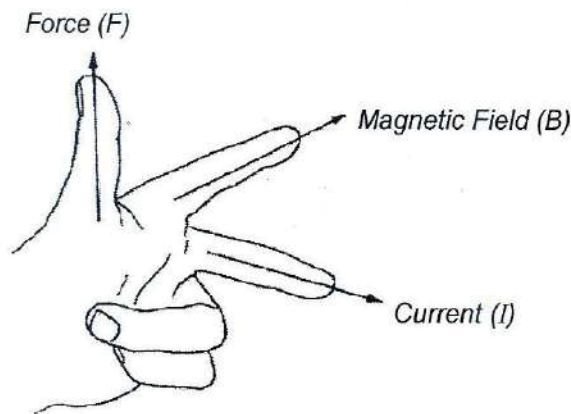


Figure 14.3.2 Fleming's Left-Hand Rule.

14.3.1 Calculating the Force on a Current in an External Magnetic Field

Fleming's Left-Hand Rule allows you to determine the direction of the magnetic force F_B on a current-carrying conductor placed perpendicular to a magnetic flux density B . However, in general, the angle between B and I can be any angle θ . The magnetic force F_B experienced by the conductor is **always perpendicular to the plane containing B and I** . Fleming's Left-Hand Rule can still be applied in such instances to determine the orientation of F_B .

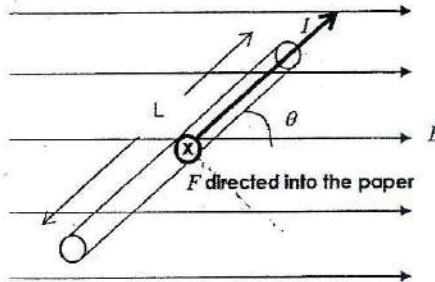
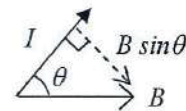


Figure 14.3.3 Magnetic Force on a Current Carrying Conductor

In the above situation, the component of the magnetic flux density perpendicular to the conductor is given by

$$B_{\perp} = B \sin \theta$$

where θ is the angle between B and I .



Hence, the magnetic force on the straight conductor is given by

$$F_B = BIL \sin \theta$$

where L is the length of the conductor carrying current, I , in the magnetic field, B and θ is the angle between B and I .

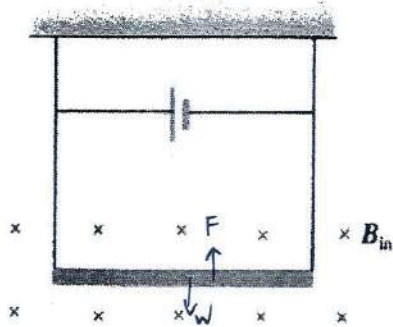
The direction of the magnetic force is given by the Fleming's Left-Hand Rule, hence directed into the page.



If you memorise the force as $F_B = BIL \sin \theta$, please remember that the **angle θ is given as the angle between the current and the B-field**. Examiners love to trick you by giving you the wrong angle in the diagram so do watch out!

Worked Example 4

A conductor suspended by two light flexible wires as shown in the figure below has a mass per unit length of 0.040 kg m^{-1} . What current must exist in the conductor in order for the tension in the supporting wires to be zero if the magnetic field in the region is 3.6 T into the page?



For tension in the supporting wires to be zero,

$$F_B = mg$$

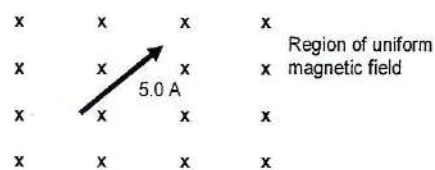
$$BIL = mg$$

$$I = 0.11 \text{ A}$$

Worked Example 5

A wire of length 3.0 cm is placed at right angles to a magnetic field of flux density 0.040 T . The wire carries a current of 5.0 A . What is the magnitude of the force which the field exerts on the wire?

- A less than 0.006 N
- ✓ B 0.006 N
- C greater than 0.006 N but less than 0.6 N
- D 0.60 N
- E greater than 0.60 N



[J93/1/19]

$$B \times I \text{ are } \perp$$

$$F_B = BIL$$

$$= 0.006 \text{ N}$$

KIASU
ExamPaper

14.4 Defining the Magnetic Flux Density and the Tesla

We may now define quantitatively the magnetic flux density, B

Definition of Magnetic Flux Density, B :

The **magnetic flux density** is defined as the **force per unit length per unit current** acting on a straight current-carrying conductor placed **perpendicular to a magnetic field**.

$$B = \frac{F}{IL \sin 90^\circ}$$

Definition of Tesla (T):

The magnetic flux density of a magnetic field is said to be 1 tesla, if the force per unit length per unit current which acts on a straight current-carrying conductor placed perpendicular to the magnetic field is 1 newton per metre per ampere.

Magnetic flux density, B , is a **vector** quantity and the **S.I. unit is the tesla (T)**.

For different shapes of current-carrying wires, the magnetic fields around the wires will in general be different. There are ways to measure experimentally the value of B , which generally depends on the current that is generating the magnetic field as well as the distance from the wire.

14.4.1 Measuring Magnetic Flux Density using a Current Balance

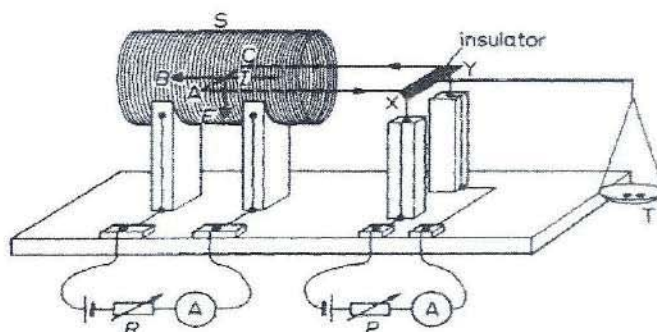


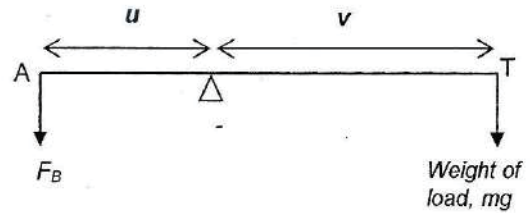
Figure 14.4.1 Current balance.

How a current balance works:

1. A wire frame AXYC is balanced on 2 knife edges through which a current I can pass from a battery. The frame is arranged such that the side AC of the frame (of length L) lies within the magnetic field whose flux density B is to be determined.
2. When there is no current, the frame is horizontal, as the force acting on AC is zero. When current flows through, a magnetic force acts on AC and pushes that side of the frame downwards (by Fleming's Left-Hand Rule).
3. A known mass is loaded on the right side to restore the frame to its horizontal position.

4. Consider the free-body diagram of the wire frame, taking moments about the knife edges,
Clockwise moments = anticlockwise moments
 $(mg) v = (BIL)u$

$$B = \frac{mgv}{ILu}$$



Side view of wire frame.

14.5 Forces Between 2 Long Parallel Current-Carrying Conductors

Now that we know that a current can generate magnetic fields like a magnet and that they can also experience a magnetic force when they are near other magnets, it is natural to consider what happens if we replace the magnet with another current-carrying wire.

When 2 parallel long wires are placed in a vacuum, each wire experiences a force as it will be in a magnetic field generated by the other wire.

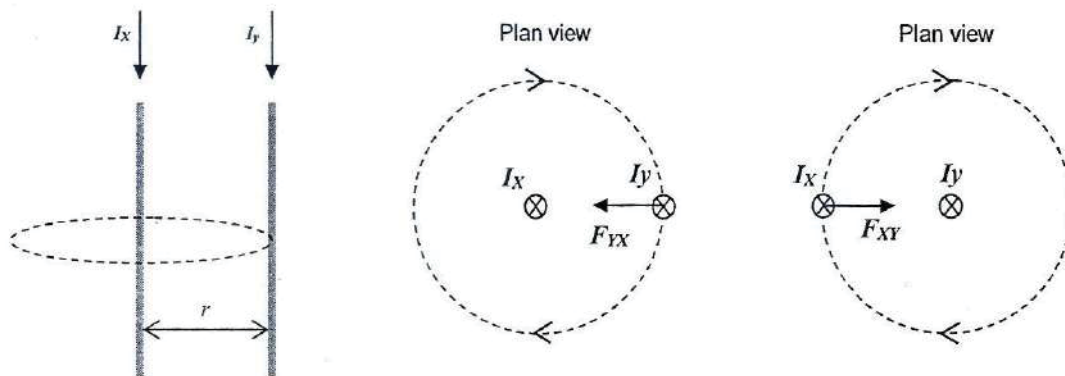


Figure 14.5.1 Two long parallel current-carrying conductors.

Current I_X sets up a magnetic field that interacts with current I_Y . Therefore the force acting on I_Y due to B_X is

$$F_{YX} = B_X I_Y l$$

Recall that B_X is given by

$$B_X = \frac{\mu_0 I_X}{2\pi r}$$

Hence the force

$$F_{YX} = \frac{\mu_0 I_X I_Y l}{2\pi r}$$

The direction of the force on I_Y can be found using Fleming's Left-Hand Rule. In this case, the force on I_Y due to I_X is towards I_X . We can go through the same analysis to understand the force acting on I_X as a result of I_Y , but honestly, there is no need to do so. This is because we can simply apply **Newton's 3rd law** and conclude that the force on I_X due to I_Y is of the same magnitude but in the opposite direction!

Hence the force on I_X due to I_Y is also $F_{XY} = \frac{\mu_0 I_X I_Y l}{2\pi r}$.

It is interesting to note that in this case where the **two currents are flowing in the same direction**, the **force between them is attractive**. We can apply a similar analysis to currents flowing in opposite directions and find that the force between them is repulsive.

A simple way of remembering the conclusion is

"LIKE CURRENTS ATTRACT, UNLIKE CURRENTS REPEL."

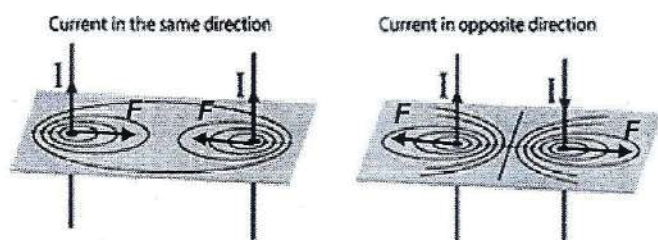


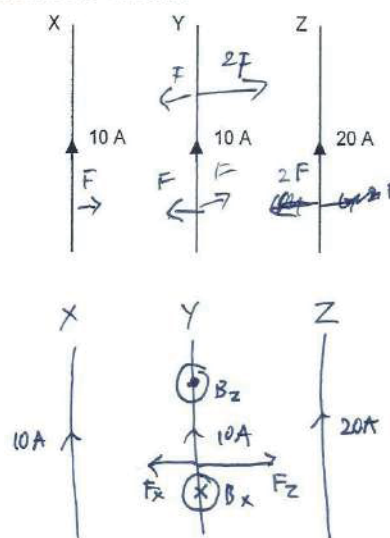
Figure 14.5.2 Sketch showing the magnetic fields and forces on parallel current-carrying wires.

Worked Example 6

X, Y and Z are three coplanar long parallel wires carrying currents as shown.

The resultant force on wire Y is

- A zero
- B towards X
- ☒ C towards Z
- D perpendicular to the paper

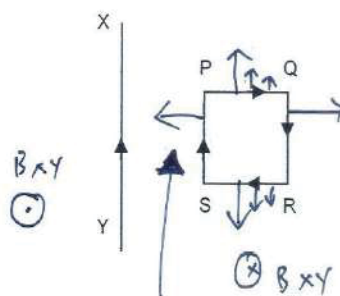


Worked Example 7

A long straight wire XY lies in the same plane as a square loop of wire PQRS which is free to move. The sides PS and QR are initially parallel to XY. The wire and loop carry steady currents as shown in the diagram.

What will be the effect on the loop?

- ☒ A It will move towards the long wire.
- B It will move away from the long wire.
- C It will rotate about the axis parallel to XY.
- D It will be unaffected.
- E It will contract.



→ coplanar, no rotation.

KIASU
ExamPaper



14.6 Revisiting the DC Motor

Consider a rectangular coil PQRS of N turns pivoted so that it can rotate freely about a vertical axis which is perpendicular to a uniform horizontal magnetic field of flux density B . Let the normal to the plane of the coil make an angle θ with the field (Figure 14.6.1 (b)).

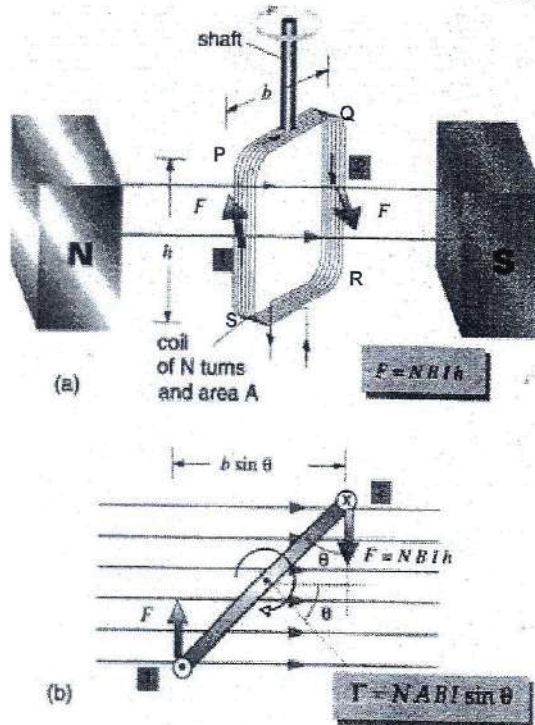


Figure 14.6.1

The DC motor.

The side PS of length h is perpendicular to the magnetic field (from N to S) and so the magnetic force on it is given by

$$F = BIh$$

If the coil has N turns, the length of the conductor is increased N times and so the force acting on side PS is

$$F = NBIh$$

The force acting on side QR is also given by $F = NBIh$, but its direction is opposite to that on PS, as shown in Figure 14.6.1 (b). The force on side PQ cancels out the force on side SR as these two forces are equal in magnitude but opposite in direction.

The two forces on sides PS and QR tend to turn the coil about an axis XY passing through the middle of the coil. These two forces together form a couple and the torque due to this couple is given by

$$\tau = F(b \sin \theta)$$

$$\tau = NBIhb \sin \theta$$

But $h \times b = \text{Area, } A \text{ of the coil. So}$

$$\tau = NBI A \sin \theta$$

14.7 Magnetic Force on a Moving Electric Charge

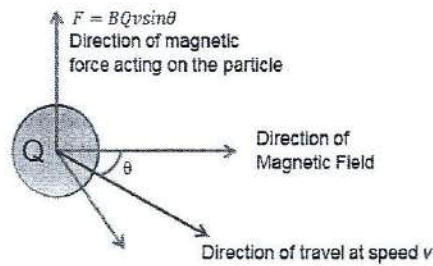


Figure 14.7.1 Magnetic Force acting on a Moving Charge

The magnitude of the force acting on a charge Q moving with velocity v in a magnetic field of flux density B is given by

$$F_B = B Q v \sin \theta$$

where θ is the angle between v and B .

Note:

1. The magnetic force is maximum when the charge is moving perpendicularly to the magnetic field.
2. A charge in a magnetic field does not experience any force when
 - It is moving parallel to the direction of the magnetic field
 - It is stationary in the field

14.8 Motion of a Charged Particle in a Uniform Magnetic Field

Let's now look at how moving charges that are not confined within a wire is affected by a magnetic field. We will again use Fleming's Left Hand Rule to find the direction of this force.

Consider a positively charged particle entering a magnetic field with velocity v .

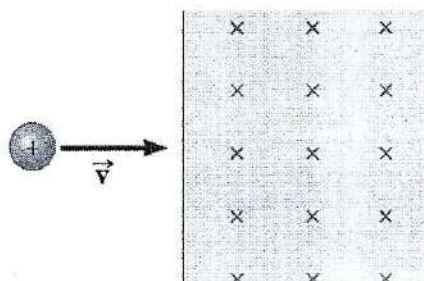


Figure 14.8.1 Charge entering a magnetic field

Using Fleming's Left Hand Rule and using the concept that the direction of the convention current is the direction of motion of the positively charged particle, we can conclude that the direction of the force at this instant is upwards. This causes the particle to be deflected from its initial path.

However, note that the direction of the force remains perpendicular to the direction of velocity and magnetic field (Fleming's Left Hand Rule) even as the charged particle deflects, hence the charged particle will move in a circular path! The magnetic force provides the centripetal acceleration needed for the charged particle to travel in a circular path.

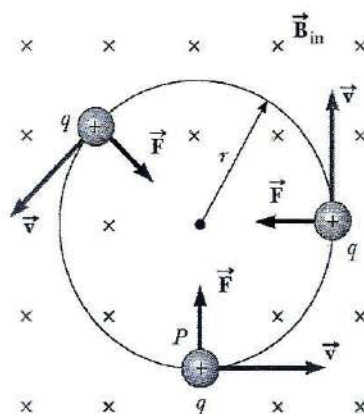


Figure 14.8.2 A positively charged particle performing circular motion in a magnetic field

The magnetic force on the positive charge provides the centripetal force:

$$Bqv = \frac{mv^2}{r}, \text{ where } r \text{ is the radius of the circular path}$$

$$r = \frac{mv}{Bq}$$

From this equation for the radius of the path, we can deduce that for particles entering the magnetic field with the **same velocity**, the **radius would be proportional to the mass-to-charge ratio** of the particle (i.e. $\frac{m}{q}$). The ratio of the charge on a particle to the mass of the particle (i.e. $\frac{q}{m}$) is known as the **specific charge**.

If T is the time taken to complete one revolution of the circular path and v is the constant speed of the charge,

$$v = r\omega$$

$$v = r \frac{2\pi}{T}$$

$$v = \left(\frac{mv}{Bq}\right) \frac{2\pi}{T}$$

$$T = \frac{2\pi m}{Bq}$$

The period T is dependent on the mass, magnetic flux density and the charge of the particle. It is independent of the speed of the particle or the radius of the circular path.

Helpful
Tips

1. Similar to the force acting on straight current carrying conductor placed in the magnetic field, we have to remember that the angle θ is the angle between the direction of motion and the magnetic field lines. This is a similar trick used by examiners.
2. For negatively charged particles moving in a magnetic field, the magnitude of the force is still the same. However, when determining the direction of the force, for *positively charged particles*, the direction of I is the direction of v , but for *negatively charged particles*, the direction of I is opposite to the direction of v .
3. It is very common in exam questions to be asked to draw the path taken by moving charges in a magnetic field. Apply Fleming's Left Hand Rule (remember the sign of the charge is important!) to find the direction of the magnetic force, which is also the direction of the centripetal force. The centripetal force acts towards the center of the circular motion, and hence deflect the moving charge in the direction of the force, or imagine where the center of the circular motion is, and you should be able to draw the path correctly. Please ensure that the path is as circular as possible, that is, you should draw with a set of compass. If you really don't have one in the exam, then label the path drawn as 'circular path'.
4. The **magnetic force does no work on the charged particle during the circular motion** as it is perpendicular to v . Hence, the speed and kinetic energy of the charge remains the same as it moves along the circular path.

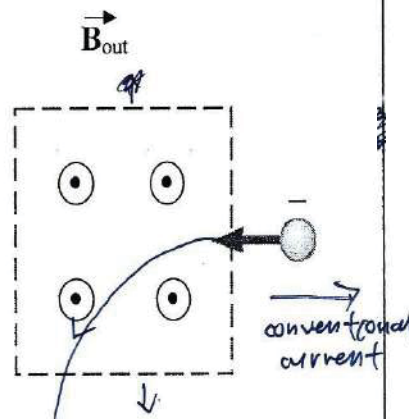
Worked Example 7

An electron with speed $2.0 \times 10^6 \text{ m s}^{-1}$ enters a magnetic field of flux density 2.0 mT perpendicular to its direction of motion. The magnetic field only exists within a square of size 8.0 mm . Draw the path taken by the electron as it moves through the magnetic field.

$$Bqv = \frac{mv^2}{r}$$

$$\Rightarrow r = \frac{mv}{Bq} = \frac{(9.11 \times 10^{-31})(2.0 \times 10^6)}{(2.0 \times 10^{-3})(1.6 \times 10^{-19})} = 5.7 \times 10^{-3} \text{ m}$$

To determine direction of deflection
Apply FLHR, direction of magnetic force acting on



14.9 Moving charges in E and B fields

It is possible for a charged particle to move into a region where both an electric field and a magnetic field are applied. The charged particle will experience an electric force and a magnetic force simultaneously. **To determine the final motion of the charged particle, we will have to go back to Newton's laws of motion which helps us to predict the motion once we know the resultant force acting on the charged particle.**

In the special case where the electric force and magnetic force are of **equal magnitude and opposite in direction**, the charged particle will travel in a straight line as the resultant force is zero.

Helpful
Tips

1. Note that we have neglected the gravitational force acting on the electron. This is a good approximation when we are dealing with atomic particles. For an electric field of 10^4 N C^{-1} , the ratio of the magnitude of the electric force qE to the magnitude of the gravitational force mg is on the order of 10^{14} for an electron and 10^{11} for a proton.
2. The electron only experiences a electric force when in the E field. Once it leaves the E field set up by the parallel plates, it will obey Newton's first law of motion and continue to move in a straight line at constant speed. When asked to draw the path, label "parabolic" within the E-field and make sure that the point where it leaves the E field, the path must be tangent to the parabolic path, then followed by a straight line.

The table below summarizes the differences in the deflection of charged particles in a magnetic and electric field.

Deflection in a magnetic field	Deflection in an electric field
1. The magnetic field can exert a magnetic force only on a moving charged particle. Stationary charged particles experience no force.	1. The electric field always exerts an electric force on a charged particle, whether it is stationary or moving.
2. The magnetic force is perpendicular to the magnetic field and the direction of motion of the charged particle. Deduced from Fleming's Left Hand Rule.	2. The electric force acts in the direction of the electric field. Deduced from the law of electrostatics (i.e. Like charges repel; unlike charges attract.)
3. Magnetic force is dependent on the speed and direction of motion of the charged particle. ($F_B = Bqv\sin\theta$)	3. Electric force is not dependent on the speed and direction of the charged particle. ($F_E = qE$)
4. Circular motion is obtained when a charged particle enters a magnetic field perpendicularly.	4. Parabolic motion is obtained when a charged particle enters a uniform electric field perpendicularly.

14.10 Velocity Selector

In many experiments involving moving charged particles, it is important that all particles move with essentially the same speed. This can be achieved by applying a combination of an electric field and a magnetic field oriented as shown in the figure below.

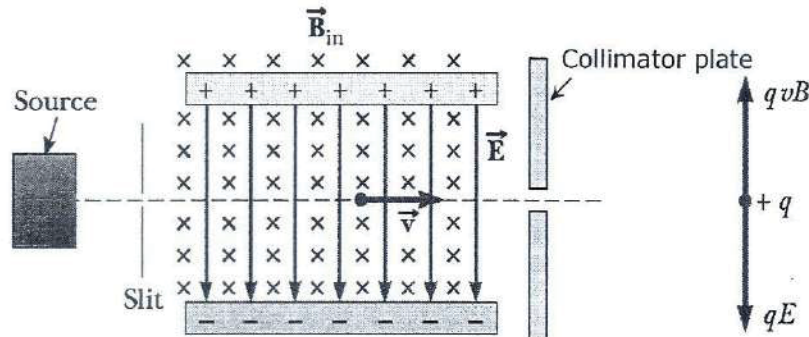


Figure 14.10 Velocity Selector

A uniform E field is applied vertically downwards and a uniform magnetic field is applied in the direction perpendicular to the electric field. Consider a positive charge entering the region from the right with velocity v .

It experiences a electric force qE that is downwards (in the same direction as the electric field) and a magnetic force Bqv that is upwards (Fleming's Left Hand Rule). By adjusting the magnitudes of the electric and magnetic fields, we can set up a scenario so that $qE = Bqv$ and the particle experiences no resultant force, and moves through the fields in a straight horizontal line!

$$\begin{aligned}\Sigma F &= 0 \\ F_B &= F_E \\ Bqv &= qE \\ v &= \frac{E}{B} = \frac{V}{Bd}\end{aligned}$$

When we have many particles of different speeds being projected into this region, only particles of velocity $v = \frac{E}{B}$ will pass through undeflected.

For particles with speeds **greater than this**, the magnetic force will be larger than the electric force and deflect upwards.

For particles with **speeds less than this**, the magnetic force will be less than the electric force, and they deflect downwards.



Not all E and B fields that are perpendicular to each other will work as a velocity selector. It is important to note that the concept behind this application is that the **particle experiences no resultant force when travelling in the crossed fields**. Hence it is important to ensure that the **direction of the electric force is opposite to the direction of the magnetic force** set up by the two fields.

APPENDIX A: Earth as a Magnet

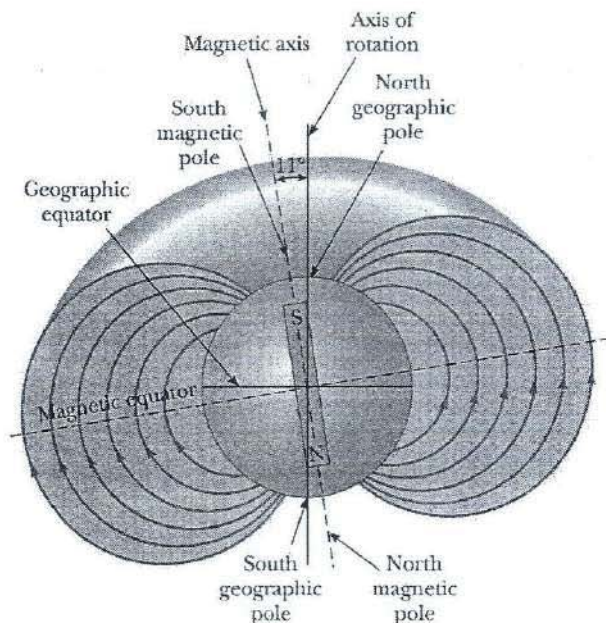
Our Earth possesses a weak magnetic field. The Earth's magnetic field pattern is as if a gigantic bar magnet is embedded within it. As shown in Figure A.1, the orientation of this fictitious bar magnet defines a magnetic axis for the Earth.

The location where the *magnetic axis* crosses the surface in the northern hemisphere is known as the *south magnetic pole*. The *north geographic pole* is that point where the Earth's *axis of rotation* crosses the surface in the northern hemisphere.

The south and north magnetic poles are found near the north and south geographic poles respectively. When we talk about a compass magnet having a north pole and a south pole, we should say more properly that it has a 'north-seeking' pole and 'south-seeking' pole. Since unlike poles attract, the *south magnetic pole* of earth's fictitious bar magnet is in fact the location toward which the 'north-seeking' end of a compass needle points.

The magnetic axis is not exactly aligned with the Earth's rotational axis. As a result, although the north pole of a compass needle points northward, it does not point exactly at the north geographic pole.

Also, Earth's magnetic field lines are not parallel to the surface at all points. Near the magnetic poles, the field lines are almost perpendicular to the surface of the Earth. At the magnetic Equator, the magnetic field lines are almost parallel to the surface of the earth. The angle that the magnetic field makes with respect to the surface at any point is known as the **angle of dip or inclination**.



Source	Magnetic Flux Density (T)
Atomic nucleus	10^{12}
Medical MRI unit	1.5
Bar magnet	10^{-2}
Surface of Sun	10^{-2}
Colour T.V.	10^{-4}
Surface of Earth	5×10^{-5}
Inside human brain (due to nerve impulses)	10^{-13}

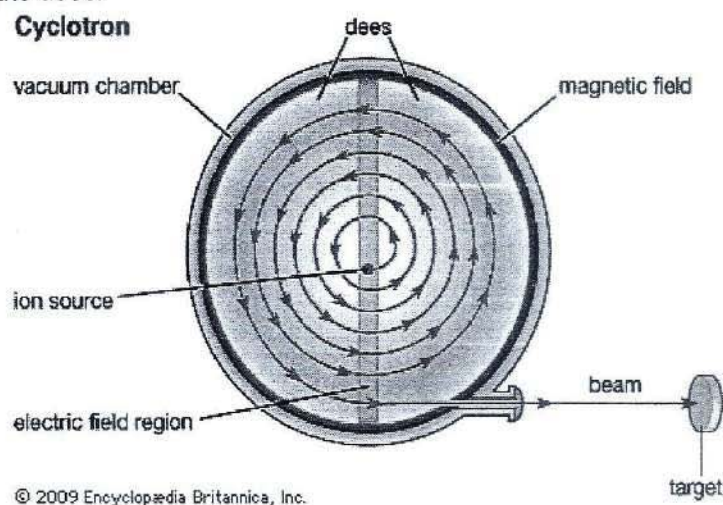
Human exposure limit is $\sim 0.2\text{T}$.

Figure A.1 Earth's magnetic poles versus geographic poles

APPENDIX B: The Cyclotron

A cyclotron is a device that can accelerate charged particles to very high speeds. The energetic particles produced are used to bombard atomic nuclei and thereby produce nuclear reactions of interests to researchers.

Both electric and magnetic forces play a key role in the cyclotron. The charges move inside two semicircular containers D_1 and D_2 , referred to as dees. A high frequency alternating potential difference is applied across the dees and a uniform magnetic field is directed perpendicularly to them. A positive ion released at P near the center of the magnet in one dee moves in a semicircular path (as shown by the dotted line) and arrives back at the gap in a time interval of $T/2$ where T is the time needed to make one complete trip around the dees.



The frequency of the applied potential difference is adjusted so that the polarity of the dees is reversed at the same time interval during which the ion travels around one dee. The ion then accelerates across the gap from D_1 to D_2 and its kinetic energy increases by qV where V is the potential difference set up across the dees.

It now moves with a larger radius in D_2 due to the increase in speed. After $T/2$, it arrives at the gap again, coinciding with the reversing of the polarity of the potential difference set up across the dees. The ion thus accelerates again across the gap and the cycle repeats itself. Every time the ion passes the gap, it gains an energy of qV . By the time it exits the cyclotron, it would have garnered about 20 MeV.

APPENDIX C : Magnetic Bottle

When charged particles move in a non-uniform magnetic field, the motion becomes complex. An example of such field is one where it is strong at the ends and weak in the middle as shown in the figure below. This configuration is known as the magnetic bottle. A charged particle starting at one end spirals along the field lines until it reaches the other end, where it reverses its path and spiral back. The charged particle is therefore trapped within it.

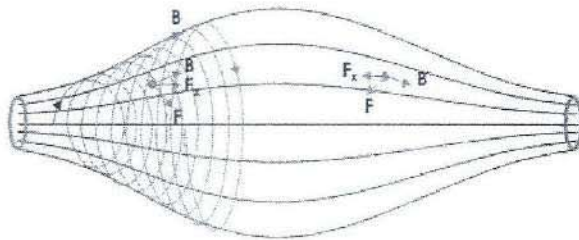
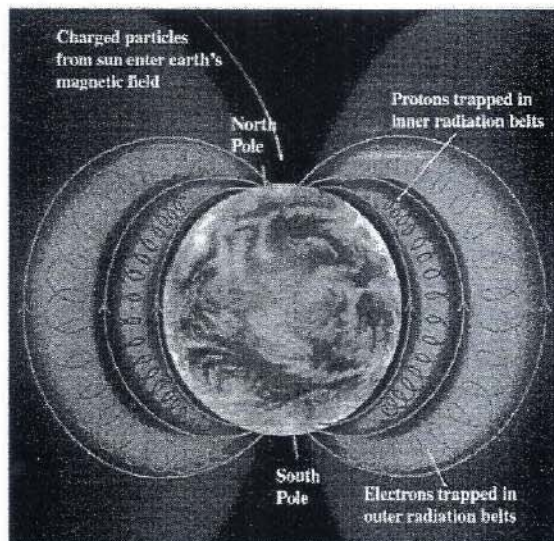


Fig 1 Magnetic Bottle

Another example is the Van Allen radiation belts consisting of charged particles (mostly electrons and protons) surrounding the Earth in doughnut-shaped regions (Figure 2).



(a)

Copyright © Addison Wesley Longman, Inc.

Figure 2 The Van Allen belts



Figure 3 The Aurora Borealis

Charged particles originate mainly from the sun (and other stars and heavenly bodies) and are commonly known as cosmic rays. Due to the Earth's non-uniform magnetic field, some of these particles are trapped and spiral around the field lines from pole to pole, making the Van Allen belts. When the particles are located near the poles, they sometimes collide with the atoms in the atmosphere, causing atoms to emit visible light. Such collisions are the origins of the beautiful Aurora Borealis (Northern Lights) and Auroras Australis.

Topic 15: Electromagnetic Induction

C2017

Electromagnetic Induction

Syllabus 9749

Content

- Magnetic flux
- Laws of electromagnetic induction

Learning Outcomes

Candidates should be able to:

- (a) define magnetic flux and the weber.
- (b) recall and solve problems using $\phi = BA$.
- (c) define magnetic flux linkage.
- (d) infer from appropriate experiments on electromagnetic induction:
 - i) that a changing magnetic flux can induce an emf,
 - ii) that the direction of the induced emf opposes the change producing it,
 - iii) the factors affecting the magnitude of the induced emf.
- (e) recall and solve problems using Faraday's law of electromagnetic induction and Lenz's law.
- (f) explain simple applications of electromagnetic induction.

15.1 Introduction

In the previous chapter, we discussed the production of magnetic fields due to electric currents. As nature is often symmetric, this led scientists to suspect that the reverse (i.e. the production of electric currents due to magnetic fields) could also be possible. Experiments conducted by Michael Faraday and independently by Joseph Henry in 1831 showed that a changing magnetic field could indeed produce an electric current in a circuit. These results led to a fundamental law known as Faraday's Law. In this chapter, we will discuss Faraday's Law and see how his experimental findings have helped set the foundation for many of modern world's technologies.



Figure 15.1.1
Michael Faraday (1791 – 1867),
British Physicist and Chemist

15.2 Magnetic Flux and Magnetic Flux Linkage

We have learnt that magnetic flux density, B , is a measure of magnetic fields. Before we learn about electromagnetic induction, we need to understand what is meant by the term magnetic flux.

Definition of Magnetic Flux, ϕ :

The **magnetic flux** through a plane surface is the product of the magnetic flux density normal to the surface and the area of the surface,

$$\phi = B_{\perp} A$$

where ϕ : magnetic flux through the plane surface

$B_{\perp}$: magnetic flux density perpendicular to the plane surface

A : area of the surface

The magnetic flux for two different cases is given below.

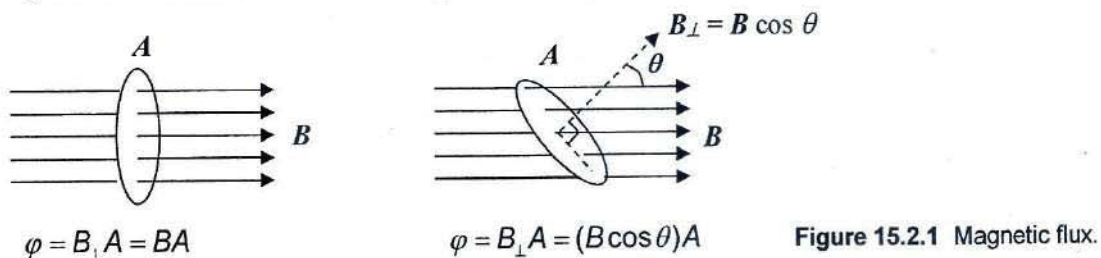


Figure 15.2.1 Magnetic flux.

The S.I. unit for magnetic flux, ϕ , is the weber (Wb).

Definition of weber (Wb):

There is a magnetic flux of 1 weber through a surface if a magnetic field of flux density of 1 T exists perpendicularly to an area of 1 m².

$$1 \text{ Wb} = 1 \text{ T} \times 1 \text{ m}^2$$

In the previous chapter, we learnt that the S.I. unit for the magnetic flux density, B , is the tesla (T). An alternative unit for B is Wb m⁻².

Note that **magnetic flux density, B , is a vector quantity, but magnetic flux, ϕ , is a scalar.**

It is common to encounter coils with multiple turns (each turn of the coil has the same area). When placed in a uniform magnetic field, every turn of the coil 'captures' the same amount of magnetic flux. The term magnetic flux linkage is commonly used to refer to the *total* magnetic flux through the coil.

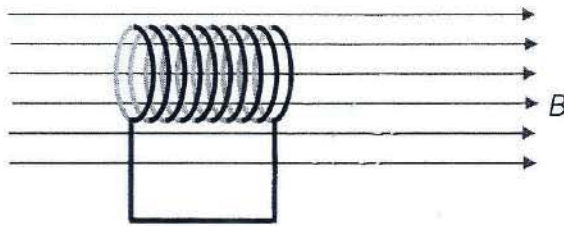


Figure 15.2.2 Magnetic flux linkage.

The **magnetic flux linkage** is defined as the product of the number of turns of the coil and the magnetic flux through each turn.

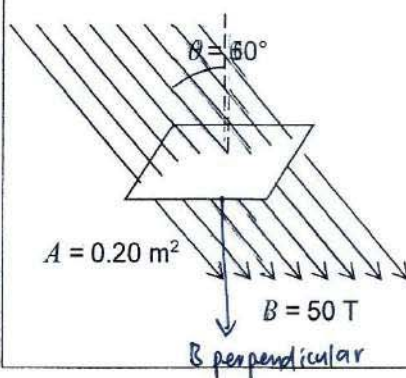
$$\Phi_{\text{linkage}} = N\phi = NB_{\perp}A$$

The S.I. unit for magnetic flux linkage is weber turns (Wb turns).

Note: In many textbooks, there is no distinction between the symbols used for magnetic flux and magnetic flux linkage, both are usually represented by Φ , hence it is important for us to know clearly in each scenario whether Φ refers to magnetic flux or magnetic flux linkage.

Example 1 Determination of magnetic flux

A flat rectangular loop of wire is placed in a uniform magnetic field which is directed at an angle to the normal of the loop (for clarity, only one row of arrows is drawn to represent the uniform magnetic field). What is the magnetic flux through the loop?

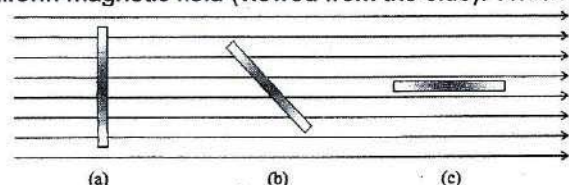


$$\begin{aligned}\phi &= B_{\perp} A \\ &= B \cos \theta A \\ &= 50 \cos 60^{\circ} \times 0.20 \\ &= 5.0 \text{ Wb}\end{aligned}$$

Example 2 Effect of orientation of magnetic field on magnetic flux

Illustrated is a loop placed at different orientation in a uniform magnetic field (viewed from the side). Which orientation gives

- 1) zero magnetic flux,
- 2) non-zero but not maximum magnetic flux and
- 3) maximum magnetic flux?



$$B_{\perp} = B$$

max
flux

$$B_{\perp} < B$$

$$B_{\perp} = 0$$

zero
flux

15.3 Induced Emf in a Moving Conductor (Motional Emf)

In the previous topic on Electromagnetism, we learnt that the magnetic force experienced by a current-carrying conductor in a magnetic field is given by $F_B = BIL \sin \theta$, where θ is the angle between B and I .

Similarly, the magnetic force acting on a moving charge Q with velocity v in a magnetic field of flux density B is given by $F_B = BQv \sin \theta$, where θ is the angle between B and v .

Next, we shall look at the effects of a moving conductor in a magnetic field.

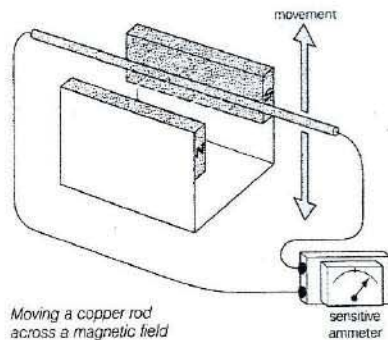


Figure 15.3.1 Conductor cutting through a magnetic field.

- When the copper rod is moved downwards, the pointer on the meter flicks to the right showing that an emf has been induced.
- When the rod is moved upwards, the meter flicks to the left. An emf has now been induced in the opposite direction.
- If the rod is held stationary in the field or made to move along the magnetic field lines, there is no induced emf.
- The **faster** the copper rod is moved, the **larger** the current.

It appears that ***an emf is induced in a moving conductor if it moves through magnetic flux. The magnitude of this induced emf is proportional to the speed.***

We shall now derive an expression for the induced emf for a moving conductor. Consider a straight conductor or moving rod of length L moving perpendicularly at a constant velocity v under the influence of some external agent through a uniform magnetic field B directed into the page as shown in Figure 15.3.2.

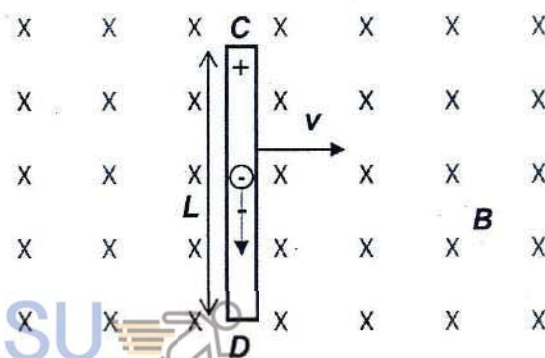


Figure 15.3.2 Isolated moving rod. The crosses indicate that the magnetic field is pointing into the page.

The electrons in the conductor experience a magnetic force $F_B = BQv$ that is directed along the length L , perpendicular to both v and B . Under the influence of this force, the electrons move to the lower end of the conductor and accumulate there, leaving a net positive charge at the upper end. As a result of this charge separation, an electric field E is produced inside the conductor. The electrons accumulate at D , leaving a

net positive charge at C, until the downward magnetic force $F_B = BQv$ is balanced by the upward electric force $F_E = QE$. At this point, the electrons stop moving. The condition for equilibrium requires that

$$\begin{aligned} BQv &= QE \\ E &= Bv \end{aligned}$$

The electric field produced in the conductor (once the electrons stop moving and E is constant) is related to the potential difference across the ends of the conductor by

$$V_{CD} = EL = BLv$$

where the upper end **C** is at a higher electric potential than the lower end **D**. Thus, from the above equation, we can see that a potential difference is maintained between the ends of the conductor as long as the conductor continues to move through the uniform magnetic field. If the direction of the motion is reversed, the polarity of the potential difference is also reversed.

In the scenario above, **only an emf is induced across the rod**. As the circuit is not complete or closed, there is *no induced current* flowing in the rod.

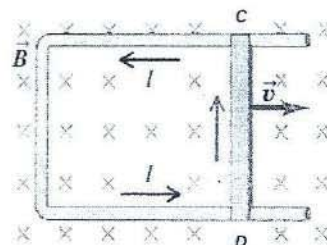


Figure 15.3.3 Rod connected to stationary conductor.

Now suppose the moving rod slides along a stationary U-shaped conductor, forming a complete circuit. No magnetic forces act on the charges in the stationary U-shaped conductor, but the charges in the moving part CD redistribute themselves, creating an electric field within it. This field establishes a current in the direction shown in Figure 15.3.3. The moving rod has become a source of electromotive force and within it, current moves from lower potential D to higher potential C, and in the remainder of the circuit, current moves from higher to lower potential. We call this emf a motional electromotive force, denoted by $\mathcal{E}$. And from the above discussion, the magnitude of this emf is

$$\mathcal{E} = BLv$$

If the total resistance of the U-shaped conductor and the sliding rod is R , the induced current I in the circuit is given by $BLv = IR$.

(i) Factors affecting magnitude of induced emf

From the formula $\mathcal{E} = BLv$, we can easily infer that increasing one or more of the following factors will increase the magnitude of the induced emf.

- The magnetic field strength B ;
- The length of the moving conductor within the magnetic field L ;
- The speed at which the conductor is moving v .

Conversely, decreasing one or more of these factors will lead to a decreased magnitude of the induced emf.

(ii) Direction of motional emf

The direction of motional emf (and induced current, if any) can be predicted by **Fleming's Right-Hand Rule**.

Often we are asked to state the polarity of the induced emf, or which end of the conductor is at the higher/lower potential. To answer that question, we must remember that we are **modelling the conductor as an emf source**, with its two ends acting like the two terminals of a battery.

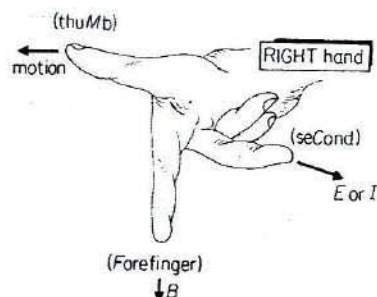
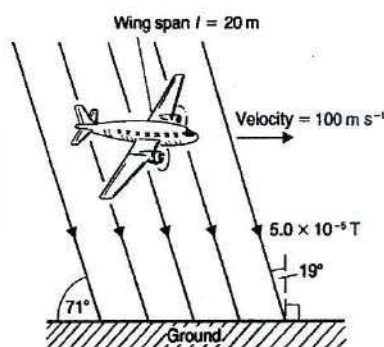


Figure 15.3.4 Fleming's Right-Hand Rule.

By modelling the conductor as an emf source, the current that will flow *within the conductor* will be from a **lower potential to a higher potential**. Hence, C is at a higher potential than D in Figure 15.3.3.

Example 3 Magnitude and polarity of induced emf

An airplane is travelling at 100 m s^{-1} in a direction which is horizontal. Calculate the emf induced between the tips of its wings, which have a span of 20 m . Take the Earth's magnetic flux density to be $5.0 \times 10^{-5} \text{ T}$ and the angle of dip at the place to be 71° . Indicate which tip is at a higher potential with a "+".



$$\begin{aligned}
 \mathcal{E} &= B_{\perp} v L = (B \cos(9^\circ)) v L \\
 &= 5.0 \times 10^{-5} \cos(9^\circ) (20) (100) \\
 &= 0.095 \text{ V}
 \end{aligned}$$

Polarity:

Use Fleming's RHR & model after a physical emf source where current flows from

15.4 Laws of Electromagnetic Induction

15.4.1 Faraday's Experiment and Findings

In the last section we saw that relative motion between a magnetic field and a conductor can induce an emf in the conductor. An experiment carried out by Michael Faraday in 1825 showed how the same effect can be achieved with no apparent relative motion. Figure 15.4.1 shows the setup that led to his discovery.

A primary coil is connected to a battery and a switch. The coil is wrapped around a wooden core, and a current in the coil produces a magnetic field when the switch is closed.

A secondary coil also is wrapped around the wooden core and is connected to a sensitive galvanometer. No battery is connected in the secondary circuit and the secondary coil is not connected to the primary coil.

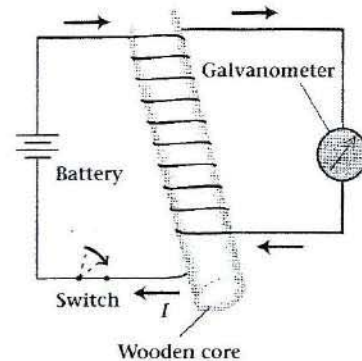


Figure 15.4.1 Faraday's experiment.

Faraday observed that:

- At the instant the switch is closed, the galvanometer needle deflects in one direction and returns to zero.
- When there is a steady current in the primary coil, the galvanometer always shows null deflection, no matter how large the current is.
- At the instant the switch is opened, the galvanometer needle deflects in the opposite direction and returns to zero.

In order to explain these observations, Faraday invoked the concept of changing magnetic flux linkage between the coils. His given explanation was:

- (1) When the switch is closed, the **magnetic field produced by the current in the primary circuit increases** from zero to a constant value over some finite time. The **secondary coil experiences this changing magnetic flux linkage produced by the primary circuit** and the induced current only exists for a short time while the magnetic flux linkage through the secondary coil is increasing.
- (2) Once the magnetic field produced by the primary coil reaches a steady value, the **magnetic flux linkage through the secondary coil is constant** and there is no current in the secondary coil.
- (3) When the switch is opened, the opposite occurs – the **magnetic field produced by the current in the primary circuit decreases** from the constant value to zero over time. Again, the **secondary coil experiences this changing magnetic flux linkage produced by the primary circuit** and the induced current flows [in the opposite direction to that in (1)] while the magnetic flux linkage through the secondary coil is decreasing.

In effect, when induced current flows in the secondary coil, the secondary circuit behaves as though a source of emf were connected to it. Hence, Faraday concluded that **there will be an induced emf in the secondary circuit if it experiences a changing magnetic flux linkage.**

With further experiments, he concluded that the **induced emf is directly proportional to the rate of change of magnetic flux linkage**, which led to the Law of electromagnetic induction.

15.4.2 Faraday's Law

The **magnitude** of the induced e.m.f. ε is **directly proportional** to the **rate of change of magnetic flux linkage** or **rate of cutting of magnetic flux**.

$$\varepsilon \propto \frac{d\Phi}{dt}$$

The S.I unit for induced emf, ε , is the volt (V). An alternative unit for induced emf is Wb turns s^{-1} .

15.4.3 Lenz's Law

The Russian physicist Heinrich Lenz (1804 – 1865) made a discovery that allows the polarity of the induced emf to be determined. This discovery led to the Lenz's Law.

Lenz's Law states that the **polarity of the induced emf** is such that it tends to produce a current that **creates a magnetic field so as to oppose the change in magnetic flux**.

Consider the following scenario:

If the bar magnet is moved rightwards towards the stationary metal loop (Figure 15.4.2), the flux through the loop increases with time. The increasing flux induces an emf in the metal loop (Faraday's Law), and since the loop forms a closed circuit, a current is produced. This induced current produces its own magnetic field. By Lenz's law, the direction of the induced current must be such as to produce a magnetic field directed in the **opposite direction** to the **strengthening external field**.

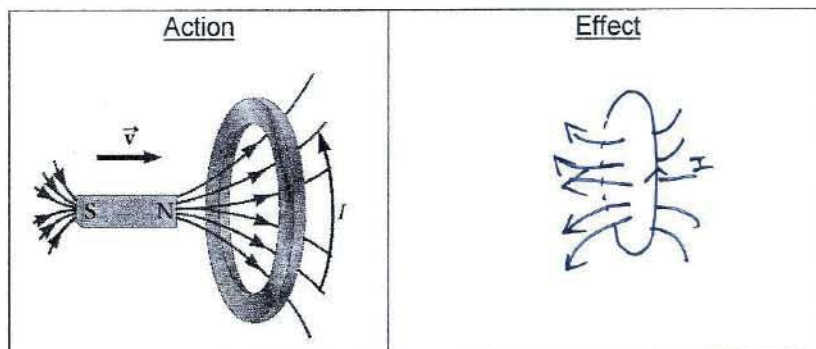


Figure 15.4.2 Magnet approaching loop.

If the magnet is moved to the left (Figure 15.4.3), the flux through the loop decreases with time. The decreasing flux again induces an emf in the metal loop (Faraday's Law). However, this time the direction of the induced current is such as to produce a magnetic field that is directed in **the same direction** as the **weakening external field**.

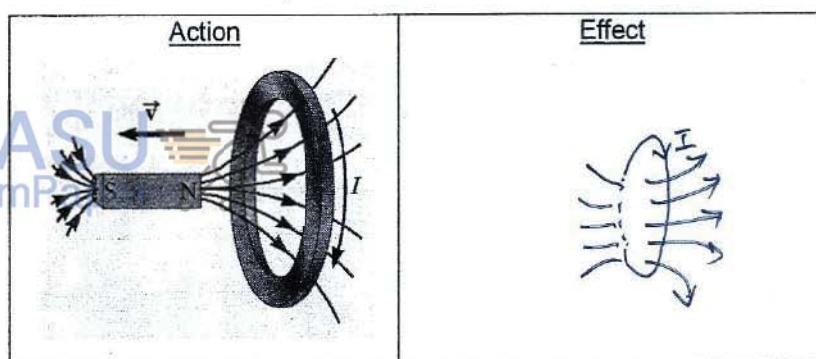


Figure 15.4.3 Magnet moving away from loop.

Notice also that the interaction of the two magnetic fields (produced by the magnet and the loop, respectively) produces magnetic forces. Note that **the direction of the induced current is such that the magnet experiences a repulsion force when approaching the loop and an attractive force when moving away from the loop.**

This leads to the important insight that **Lenz's Law is really a re-statement of the law of conservation of energy.** The direction of induced current is always such as to (produce a magnetic force to) oppose the motion of the magnet (which was the cause of the electromagnetic induction). Since the motion of the magnet is opposed, work must be done in order to maintain the motion of the magnet. This mechanical work enables the conversion of energy into the electrical energy of the current. This is further illustrated in Figure 15.4.4.

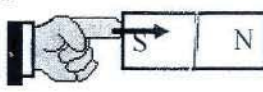
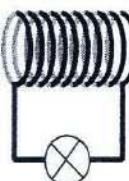
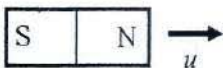
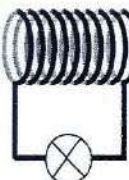
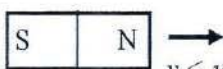
I 		An initial force is applied to the magnet to start its motion.
II 		The magnet starts to move with an initial speed u towards the coil. Hence, the magnet has a certain amount of kinetic energy.
III 		As the magnet approaches the coil, an induced emf is set up in the coil. As the coil is part of a closed circuit, an induced current will flow in a way that tends to oppose the increase in magnetic flux. Hence, the magnet slows down while the bulb lights up. Thus, the kinetic energy of the magnet is first converted into electrical energy in the circuit and in turn into light and thermal energy in the bulb.

Figure 15.4.4 Illustration of Lenz's Law in view of the law of conservation of energy.

Hence, it is common to see the following equation used in textbooks when stating Faraday's Law:

$$\varepsilon = -\frac{d\Phi}{dt}$$

The **negative sign** in the above equation is a **consequence of Lenz's law**. It implies that **the direction of the induced emf is always such as to oppose the change in the magnetic flux.**



- (1) **Faraday's Law** enables us to determine the **magnitude** of induced emf (applicable to both open and closed circuits).
- (2) **Lenz's Law** allows us to determine the **polarity (or direction)** of induced emf (applicable to both open and closed circuits) or direction of induced current (applicable only to closed circuits. If circuit is open, we usually think in terms of what would happen if it was closed and hence find the direction of the induced emf.)

Example 4 Determination of average induced emf

A flat circular coil of diameter 30 mm has 500 turns and is situated so that the plane of the coil is perpendicular to a uniform magnetic field of flux density 20 mT. The flux density is reduced to zero and then increased to 20 mT in the opposite direction at a constant rate. The time taken for the whole operation is 60 ms.

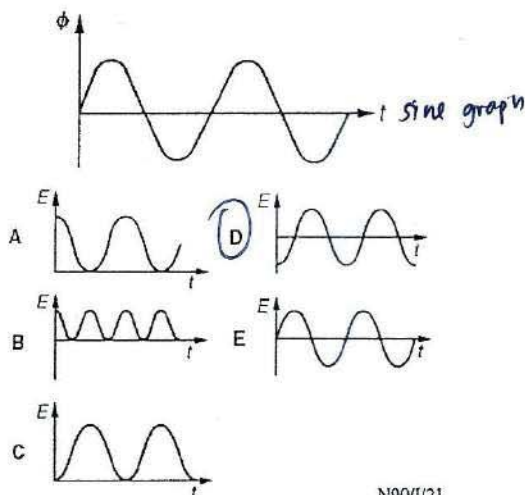
What is the average value of the emf induced in the coil?

- A 0
B 0.12 V
C 0.24 V
D 0.94 V

$$\langle \mathcal{E} \rangle = \frac{\Delta \phi}{\Delta t} = \frac{\Delta (NBA)}{\Delta t} = NA \frac{\Delta B}{\Delta t} = 500 \left(\frac{\pi (30 \times 10^{-3})^2}{4} \right) \left(\frac{20 - (-20)}{60} \right) = 0.24 \text{ V}$$

**Example 5** Graphs of varying magnetic flux and emf induced

The magnetic flux Φ through a coil varies with time as shown in the diagram. Which graph best represents the variation with time of the induced emf $\mathcal{E}$ in the coil?



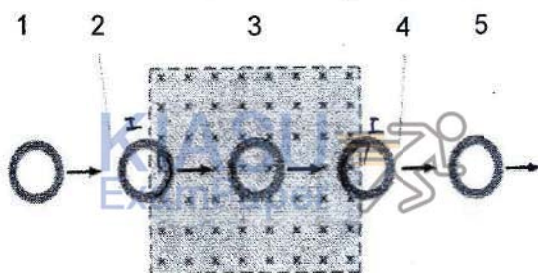
$$\mathcal{E} = -\frac{d\Phi}{dt} \rightarrow \text{gradient of flux vs } t \text{ graph}$$

-cosine graph

N90/11/21

Example 6 Application of Lenz's Law

In the figure below, there is a constant magnetic field in a rectangular region of space. This field is directed perpendicularly into the plane of the paper. Outside this region there is no magnetic field. A copper ring slides through the region, from position 1 to position 5. For each of the five positions, determine if an induced current exists in the ring and, if so, find the direction of the current.



3: Flux thru ring remains a constant.
By Faraday's law, no induced emf & current.

4: opp. of 2.

5: Flux thru ring remains a constant at zero. By Faraday's law, no induced emf & current

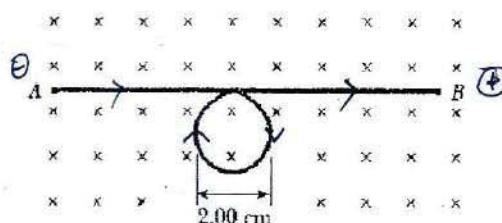
2: as the ring enters the field, the magnetic flux thru ring increases. By Faraday's law, an emf is induced in the ring. By Lenz's Law, an anticlockwise current flows in the ring to produce a magnetic field in the opp. direction to oppose the increasing flux.

Example 7 Faraday's Law and Lenz's Law

The magnetic field shown in the figure has a uniform magnitude of 25.0 mT and is directed into the paper. The initial diameter of the kink is 2.00 cm.

- (a) The insulated wire is quickly pulled taut, such that the diameter of the kink becomes zero in 50.0 ms. Determine the average induced emf between endpoints A and B. Specify the polarity of the emf in your answer. Suppose the kink is undisturbed, but the magnetic field increases to 100 mT in 4.00×10^{-3} s. Determine the average induced emf between endpoints A and B, together with the polarity.

$$\begin{aligned} \langle \mathcal{E} \rangle &= \frac{\Delta \phi}{\Delta t} = B \frac{\Delta A}{\Delta t} \\ &= \frac{(25.0 \times 10^{-3}) \left(\frac{\pi (2.0 \times 10^{-2})^2}{4} - 0 \right)}{50.0 \times 10^{-3}} \\ &= 1.57 \times 10^{-4} \text{ V} \end{aligned}$$



As the wire is quickly pulled taut, the area of loop decreases, magnetic flux thru loop decreases.

⇒ By Lenz's law, induced current flows in the clockwise direction from A through the loop to B to produce a B-field that is in the same direction to oppose the decrease in flux

⇒ polarity of induced emf: B is higher than A

(model after the inside of a physical emf source)

current flows from
* lower to higher potential for battery

⇒ opp. of what current flows normally.

15.5 Practical Applications of Electromagnetic Induction

15.5.1 Electric Generator (rotating coil in magnetic field)

A coil of N turns, with area A , rotates at a constant angular velocity ω in a uniform magnetic field B .

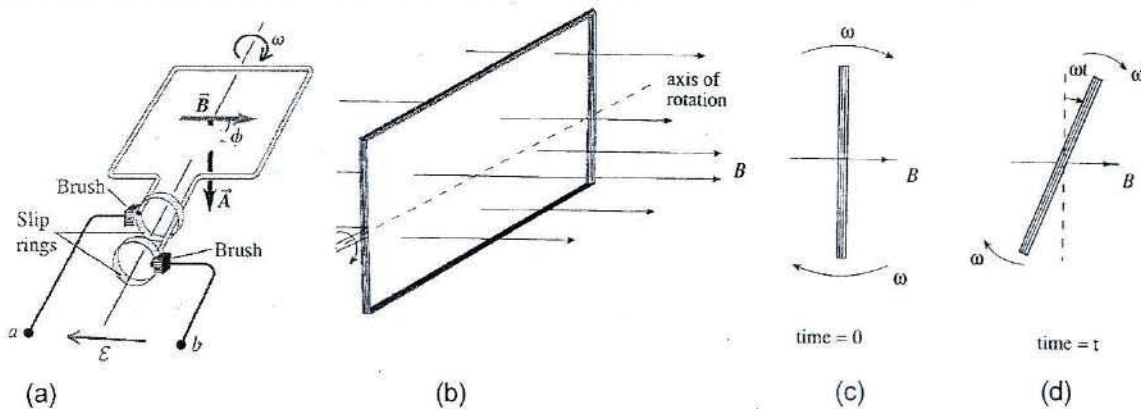


Figure 15.5.1 Rotating coil in a magnetic field, generating a current.

As the coil rotates in the magnetic field, it experiences a **changing magnetic flux linkage** as the component of the magnetic field perpendicular to the plane of the coil keeps changing. By Faraday's law, an emf is induced in the coil.

Derivation of an expression for the induced emf

1. Magnetic flux linkage through the coil

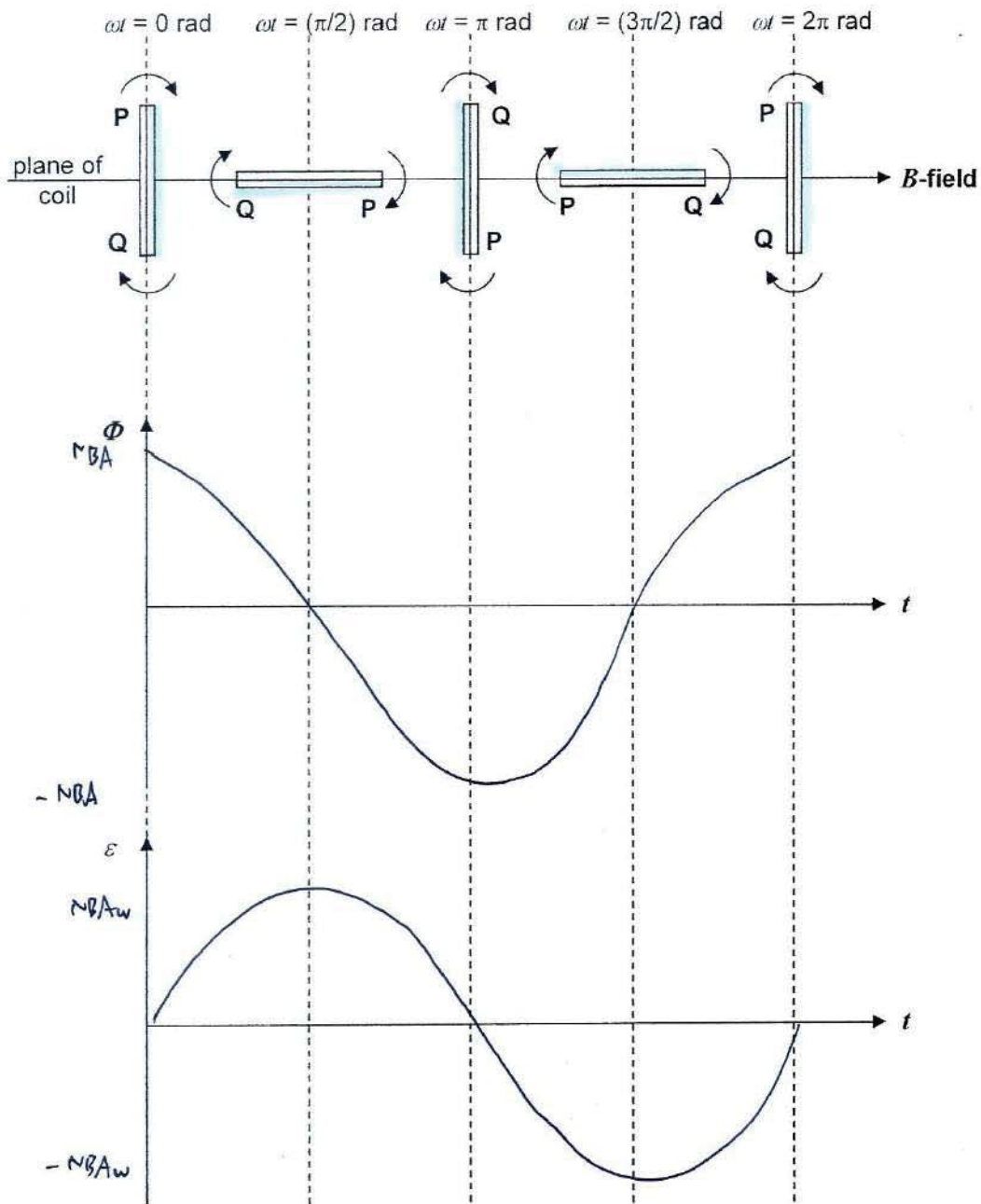
$$\Phi = NB_{\perp} A = NBA \cos \theta = NBA \cos \omega t$$

2. Induced emf

$$\begin{aligned} \varepsilon &= -\frac{d\Phi}{dt} = -\frac{d}{dt}(NBA \cos \omega t) = -NBA \frac{d}{dt}(\cos \omega t) \\ \varepsilon &= NBA\omega \sin \omega t \end{aligned}$$

Hence the induced emf is dependent on N , B , A and ω . By Lenz's law, the direction of the induced emf is such that it sets up a current to oppose the rotation of the coil.



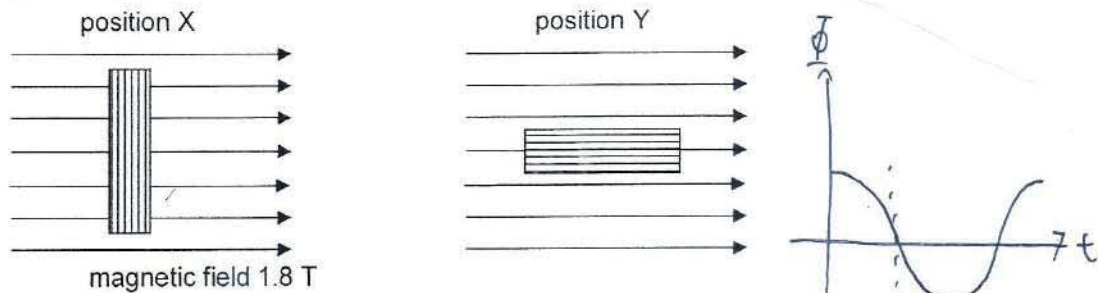


Note that the magnitude of the induced emf is largest when the magnetic flux linkage is zero. Conversely, when the magnitude of the magnetic flux linkage is largest, there is no induced emf in the coil.



Example 8 Determination of emf induced in rotating coil (N07/I/31)

A circular coil of diameter 0.020 m has 3000 turns. It is rotated in a magnetic field of 1.8 T from position X to position Y in 0.060 s.



- (a) What is the average emf induced in the coil during the rotation X to Y?
 (b) What is the average emf induced in the coil for one complete revolution, assuming constant angular velocity?

What is the instantaneous E.M.F, E at a) position X? b) position Y?

a) At max or min pts of Φ - t graph,
 $E = \frac{d\Phi}{dt}$ is 0
 At X, $E=0$

b) In b/w stationary or turning pts of Φ - t graph

$$E = \left| \frac{d\Phi}{dt} \right| \text{ is max} = NBA\omega = NB(\pi r^2) \left(\frac{2\pi}{T} \right) = 3000(1.8) \left[\pi \left(\frac{0.020}{2} \right)^2 \right] \left[\frac{2\pi}{4 \times 0.060} \right]$$

15.5.2 Speedometer of a Car

A permanent magnet is attached to the wheel of the car. A sensor consisting of a coil of wire is connected to the frame of the car. When the car moves, the magnet rotates with the wheel. As the magnet passes over the sensor, the magnetic flux in the coil rises and then falls. This induces an emf and a current in the coil. The current, in the form of pulses, is sent to an electronic device which counts the number of turns performed by the wheel in each second. With this information, the speed of the car can be calculated if the radius of the wheel is known and the speed can thus be displayed on the speedometer.

KIASU
ExamPaper



Figure 15.5.2 Speedometer.

15.5.3 Ground Fault Interrupters

The ground fault interrupter is a safety device that protects people against an electric shock when they touch appliances and power tools. In the diagram below, wire 1 leads from the wall outlet to the appliance to be protected and wire 2 leads from the appliance back to the wall outlet. An iron ring surrounds the two wires to confine the magnetic field set up by each wire. A sensing coil, which can activate a circuit breaker when changes in magnetic flux occur, is wrapped around part of the iron ring.

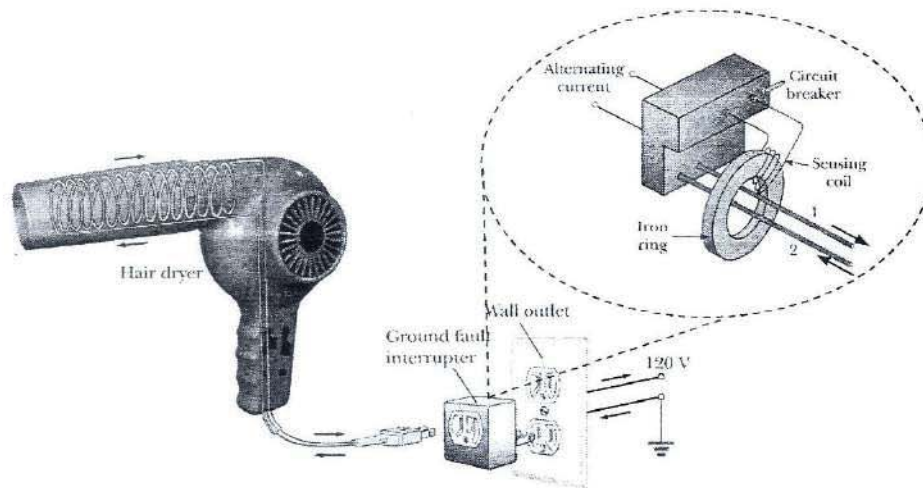


Figure 15.5.3 How the ground fault interrupter works.

Since the currents in the wires are in opposite directions, the net magnetic field through the sensing coil due to the currents is zero. However, if a short circuit occurs in the appliance so that there is no returning current, the net magnetic field through the sensing coil is no longer zero. This can happen for example if one of the wires loses its insulation, providing a path through the person to ground if the person happens to be touching the appliance. As the current is alternating, the magnetic flux through the sensing coil changes with time, thus inducing an emf in the coil. This induced emf will trigger the circuit breaker, stopping the current quickly in about a millisecond before it reaches a level that might be harmful to the person using the appliance.

15.5.4 Electric Guitar

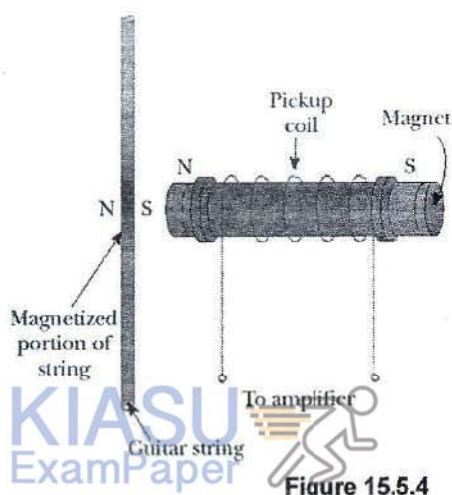


Figure 15.5.4



How the electric guitar works.

The pickup coil is placed near the vibrating guitar string, which is made of a metal that can be magnetized. The permanent magnet inside the coil magnetizes the portion of the string nearest the coil. When the guitar string vibrates at some frequency, its magnetized segment produces a changing magnetic flux through the pickup coil. The changing magnetic flux induces an emf in the coil. This induced emf or voltage is fed to an amplifier, the output of which is sent to the loudspeakers.

15.6 Eddy Currents

As we have seen, emf and current can be induced in a circuit by a changing magnetic flux. This also results in eddy currents being induced in bulk pieces of metal moving through a magnetic field.

Consider a flat copper or aluminium plate attached to a rigid rod which swings back and forth through a magnetic field.

As the plate enters the field, the changing magnetic flux induces an emf in the plate, which in turn causes the free electrons in the plate to move, producing swirling currents (eddy currents). As the currents flow, the plate will heat up: its internal energy will increase. By the principle of conservation of energy, this means that the mechanical energy in the oscillating plate must reduce. This loss in mechanical energy will eventually bring the plate to rest.

Also, according to Lenz's law, the eddy currents give rise to repulsive forces that opposes the motion of the plate. This retarding magnetic force will eventually bring the plate to a rest.

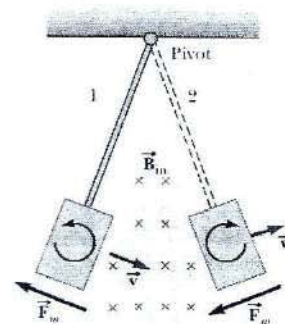


Figure 15.6.1 Eddy currents.

Eddy currents are often undesirable because they represent a transformation of mechanical energy to internal energy. If slots are cut into the plate, the eddy currents and the corresponding retarding force will be greatly reduced. Also, conducting parts in devices are often laminated, that is, they are built up in thin layers separated by a non-conducting material. This layered structure greatly increases the resistance of eddy current paths and effectively confines the currents to individual layers. Such a laminated structure is used in transformer cores.

However, eddy currents are often used in the braking system of many subway and rapid-transit cars. An electromagnet attached to the train is positioned near the steel rails. The braking action occurs when a large current is passed through the electromagnet. The motion of the magnet relative to the rails induces eddy currents in the rails, and the direction of these currents produces a drag force on the moving train. As the train slows down, the eddy currents decrease, hence the braking effect is quite smooth.

Example 9 Magnetic Braking

A copper disc spins freely between the poles of an unconnected electromagnet as shown. Describe and explain what will happen to the speed of rotation of the disc when a direct current is switched on in the electromagnet.

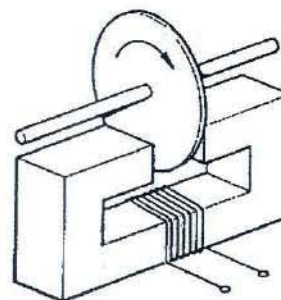
Answer:

As the copper disc rotates, different parts of the copper disc pass through the magnetic field exerted by the electromagnet.

Regions of the copper disc which are entering the electromagnet's magnetic field experience strengthening flux, leading to induced emf proportional to the rate of flux change (by Faraday's law). The polarity of the induced emf and current is such that it creates a magnetic flux in a direction opposite to the electromagnet's magnetic field, creating a repelling pole (by Lenz's law).

Regions of the copper disc leaving the electromagnet's magnetic field experience weakening flux, leading to an induced emf proportional to the rate of flux change (by Faraday's Law). The polarity of the induced emf and current is such that it creates a magnetic flux in the same direction as the electromagnet's magnetic field, creating an attracting pole (by Lenz's law).

For both regions, the direction of the magnetic force is always opposite to the direction of motion. The magnetic force retards the motion and slows down the spinning.



Appendix 1 A Faraday's Disc with area A rotating with a frequency f in a magnetic field B

The Faraday's disc, an example shown in Figure 1A, is the first homopolar generator. A homopolar generator is a DC electrical generator comprising an electrically conductive disc or cylinder rotating in a plane perpendicular to a uniform static magnetic field.

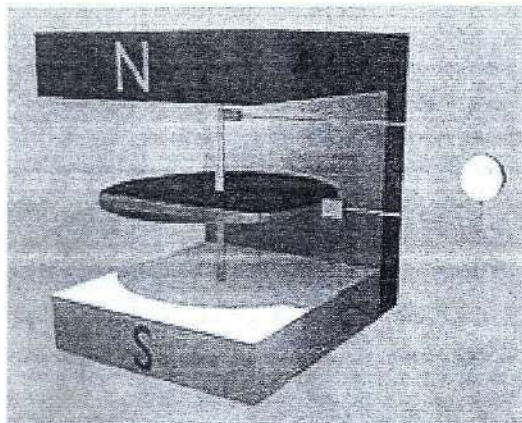


Figure 1A

As shown in Figure 1B, the disc can be divided into small elemental strips. As each strip rotates, they cut the magnetic field B . By Faraday's Law of Electromagnetic induction, an induced emf will be setup across each strip.

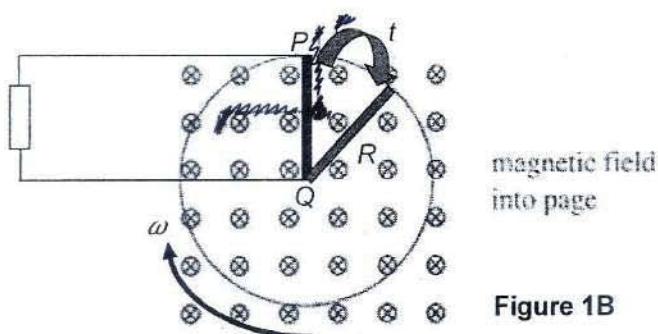


Figure 1B

Derivation of an expression for the induced emf:

Consider an elemental strip

Magnetic flux it cuts in time t ,

$$\Phi = B_{\perp} A = B \left(\frac{1}{2} R^2 \theta \right)$$

The rate at which magnetic flux is cut by the conductor is given by

$$\varepsilon = -\frac{d\Phi}{dt} = -\frac{d}{dt} \left(\frac{1}{2} B R^2 \theta \right) = -\frac{1}{2} B R^2 \frac{d\theta}{dt} = -\frac{1}{2} B R^2 \omega$$

Since $\omega = 2\pi f$ and $A = \pi R^2$,

$$\varepsilon = -\frac{1}{2} B R^2 \omega = B A f$$

Direction of the induced emf:

Electrons in the disc are pushed to the ^{centre} outer edge (circumference). The flow of electrons to the ^{centre} edge will stop when the electric force equals the magnetic force. Using Fleming's Right Hand Rule, P is at a higher potential than Q.

Appendix 2 Electromagnetic Braking in Transport.

The link below addresses some of the problems that have been encountered in using electromagnetic brakes on high-speed trains, the two main issues being the brakes interfering with signalling equipment and the excessive heating caused by the brakes in the rails.

<http://www.railwaygazette.com/news/single-view/view/eddy-current-braking-a-long-road-to-success.html>

Appendix 3 Volcanic Eruption as a Consequence of Eddy Currents in Action

(Extracted from Young and Freedman, University Physics, 12th edition)

A particularly dramatic example of eddy currents in action is Jupiter's moon Io, which is slightly larger than the earth's moon. Io moves rapidly through Jupiter's intense magnetic field (Figure 3A), and this sets up strong eddy currents within Io's interior. These currents dissipate energy at a rate of 10^{12} W, equivalent to setting off a one kilo-ton nuclear weapon inside Io every 4 seconds! This dissipated energy helps to keep Io's interior hot and so helps to cause volcanic eruptions on its surface, like those in Figure B. (However, note that gravitational effects from Jupiter cause even more heating.)

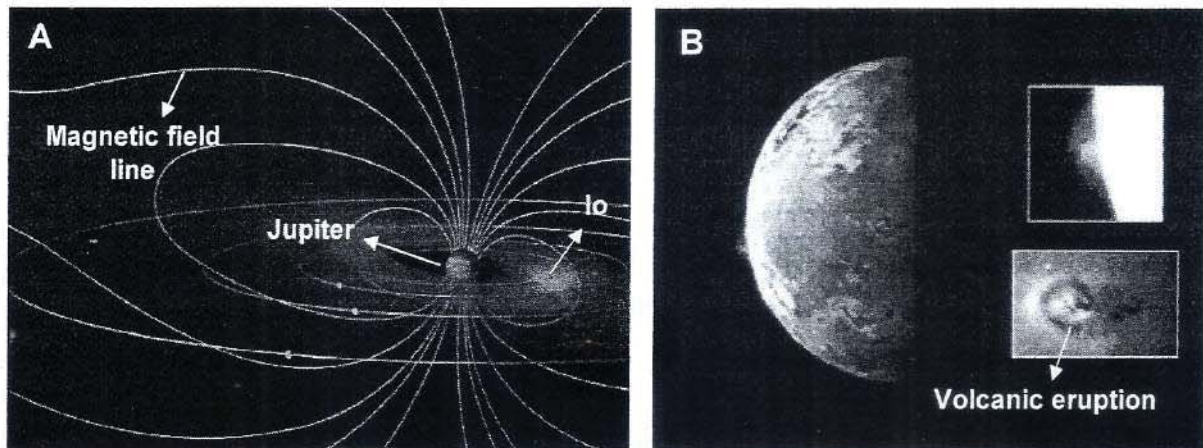


Figure 3 A Powerful magnetic field lines of Jupiter induce eddy currents in its moon Io, as Io moves in its orbit around Jupiter.
 B Close-up view of a volcanic eruption, triggered in part by eddy-current heating.



16 Alternating Current

Content (H2 Physics 9749)

- Characteristics of alternating currents
- The transformer
- Rectification with a diode

Learning Outcomes

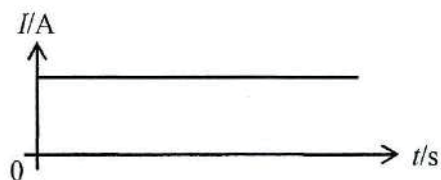
Candidates should be able to:

- show an understanding and use the terms period, frequency, peak value and root-mean-square (r.m.s.) value as applied to an alternating current or voltage.
- deduce that the mean power in a resistive load is half the maximum (peak) power for a sinusoidal alternating current.
- represent an alternating current or an alternating voltage by an equation of the form $x = x_0 \sin \omega t$.
- distinguish between r.m.s. and peak values and recall and solve problems using the relationship $I_{rms} = \frac{I_0}{\sqrt{2}}$ for the sinusoidal case.
- show an understanding of the principle of operation of a simple iron-cored transformer and recall and solve problems using $\frac{N_s}{N_p} = \frac{V_s}{V_p} = \frac{I_p}{I_s}$ for an ideal transformer.
- explain the use of a single diode for the half-wave rectification of an alternating current.

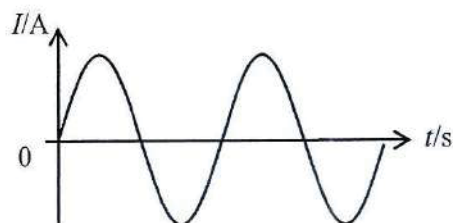
16.1 What is alternating current (AC)?

Direct current (DC) refers to current that does not change in direction. One of the most common DC sources is a battery.

On the other hand, the direction of an alternating current (AC) will vary with time. A common form of AC (e.g. the 230 V, 50 Hz AC power sockets in our homes) is one that varies sinusoidally.



Constant DC

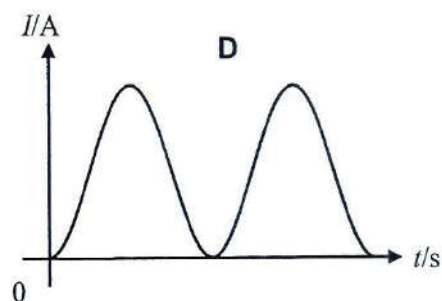
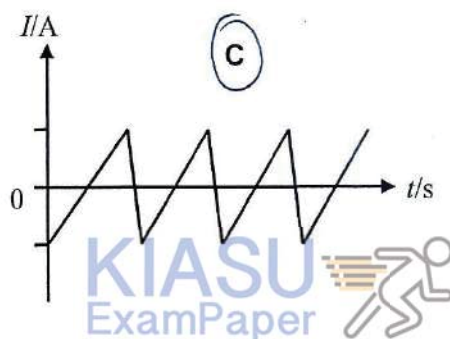
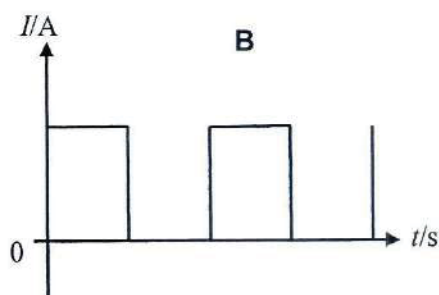
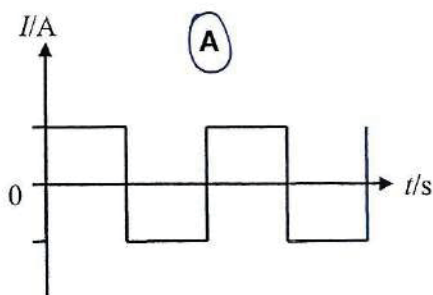


Sinusoidal AC

Alternating current occurs when charge carriers periodically reverse their direction of motion.

Example 1:

Which of the following is/are alternating current(s)?



16.2 Sinusoidal AC

The electricity supply generated for consumers' use is a sinusoidal AC.

The time-variation of a sinusoidal current or voltage can be represented by the equation

$$x = \pm x_0 \sin \omega t \quad \text{or} \quad x = \pm x_0 \cos \omega t$$

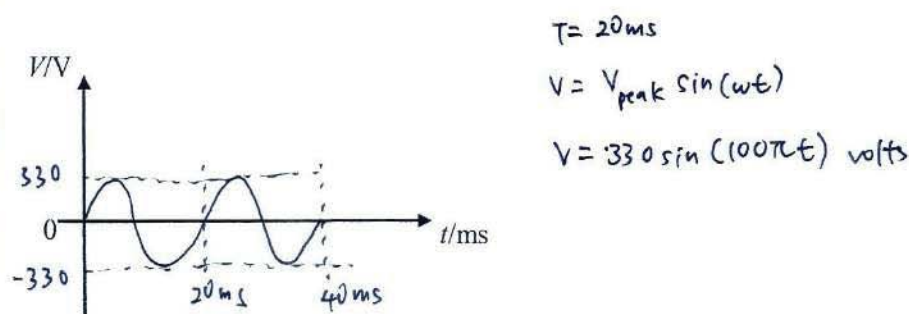
where x_0 = peak (or maximum) value

ω = angular frequency of the supply (rad s^{-1}).

Note that $\omega = 2\pi f$ where f is the frequency of the supply in hertz (Hz). Frequency is also related to the period T through the expression $f = \frac{1}{T}$.

Example 2:

The AC supply found in Singapore homes have peak voltage of 330 V and frequency of 50 Hz. Write down an expression for the voltage and sketch the waveform.



16.3 Root-mean-square (r.m.s.) values

If a steady DC passes through a resistive load R , the power P dissipated can be calculated by $P = VI$, $P = I^2 R$, or $P = \frac{V^2}{R}$. Since V and I are unchanging for a DC, there is no difference between instantaneous and average power dissipated in the resistive load.

If the current through the load is AC, the **instantaneous power** dissipated must be changing with time, since V and I are changing with time. In practice, we are often only interested in the **average power** dissipation. Interestingly, the average power can be calculated by the same set of equations

($P = VI$, $P = I^2 R$ or $P = \frac{V^2}{R}$), as long as we use the **root-mean-square** values V_{rms} and I_{rms} .

Suppose an AC periodic current I passes through an electrical load of resistance R , its instantaneous power P dissipated will be changing with time t .

$$\begin{aligned}
 \text{In general, mean power } \langle P \rangle &= \frac{1}{T} \int_0^T P \, dt \\
 &= \frac{1}{T} \int_0^T I^2 R \, dt \\
 &= R \left(\frac{1}{T} \int_0^T I^2 \, dt \right) \quad \text{which is } R \times \frac{(\text{area under } I^2 \text{ vs } t \text{ graph})}{\text{duration of one period}} \\
 &= R \langle I^2 \rangle \quad \text{which is } R \times (\text{mean-squared of } I) \\
 &= R (I_{\text{rms}})^2 \quad \text{which is } R \times [(\text{root-mean-squared of } I)]^2
 \end{aligned}$$

The **root-mean-square** value of an alternating current is that **equivalent constant direct current** that will **dissipate the same average power in a given resistive load**.

In other words, a direct current of 1 A and an alternating current of r.m.s. value of 1 A have the same heating effect, i.e. both currents dissipate the same power in a given resistor.

$$\text{Instantaneous power} = VI = I^2 R = \frac{V^2}{R}$$

$$\text{Peak power} = V_0 I_0 = I_0^2 R = \frac{V_0^2}{R}$$

$$\text{Mean power } \langle P \rangle = V_{\text{rms}} I_{\text{rms}} = I_{\text{rms}}^2 R = \frac{V_{\text{rms}}^2}{R} \quad [1]$$

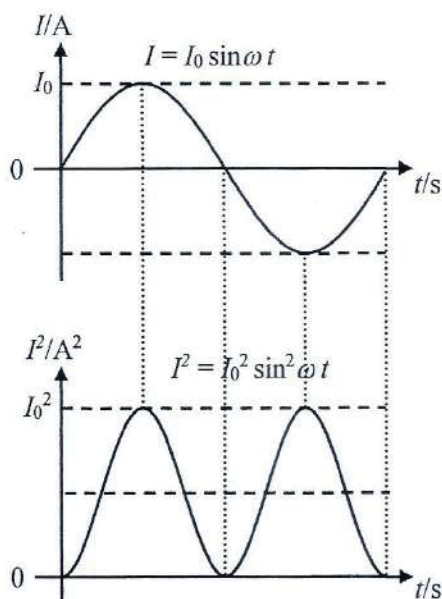
Based on the above equation in [1], we can derive that

$$V_{\text{rms}} = I_{\text{rms}} R$$

Note that the above relationships are not restricted to sinusoidal AC. They are applicable to non-sinusoidal AC as well.

16.3.1 Root-mean-square (r.m.s.) values for sinusoidal AC (IMPORTANT!)

Suppose a sinusoidal current of $I = I_0 \sin \omega t$ passes through an electrical load of resistance R .



Using concept of power dissipation,

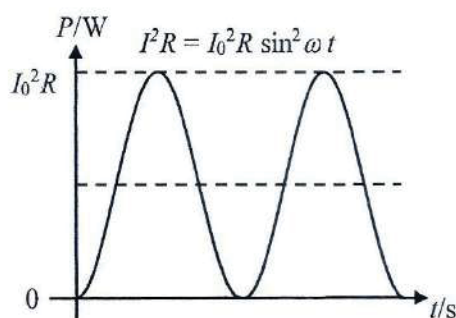
$$\text{mean power } \langle P \rangle = \frac{(\text{area under } P \text{ vs } t \text{ graph})}{\text{duration of one period}}$$

Because of the symmetrical nature of the function,

$$\begin{aligned} \langle P \rangle &= \frac{1}{2} (I_0^2 R) \quad \text{which is equal to } \frac{1}{2} P_{\max} \\ &= \left(\frac{I_0}{\sqrt{2}} \right)^2 R \end{aligned}$$

From the previous page in equation [1] which shows that

$$\begin{aligned} \langle P \rangle &= (I_{\text{rms}})^2 R, \\ \text{we can deduce that } I_{\text{rms}} &= \frac{I_0}{\sqrt{2}} \\ &\text{for a sinusoidal current.} \end{aligned}$$



In summary, for a **sinusoidal** AC,

$$I_{\text{rms}} = \frac{I_0}{\sqrt{2}} = 0.707 I_0,$$

$$V_{\text{rms}} = \frac{V_0}{\sqrt{2}} = 0.707 V_0 \quad \text{and}$$

$$\langle P \rangle = \frac{1}{2} P_{\max}$$



Note that the above 3 formulae are **ONLY applicable to sinusoidal AC**. For other forms of AC, you would need to derive the corresponding relationship between its r.m.s. value and its peak value.



For the A-level exam, AC supply is assumed sinusoidal unless otherwise stated. The value of the AC power quoted in an exam question is the mean value.

Current and voltage values are often quoted in r.m.s. values unless otherwise stated. r.m.s. values and mean power are more meaningful and useful rather than the peak values (which are just instantaneous values).

Example 3:

A $3.0\ \Omega$ resistor is connected to a sinusoidal AC supply of peak voltage 5.0 V . Calculate

- the r.m.s. voltage;
- the r.m.s. current;
- the mean power;
- a D.C. supply is used instead. What must the D.C. voltage be to produce the same power in the $3.0\ \Omega$ resistor?

a) $V_{\text{rms}} = \frac{V_{\text{peak}}}{\sqrt{2}} = \frac{5.0}{\sqrt{2}} = 3.5\text{ V}$

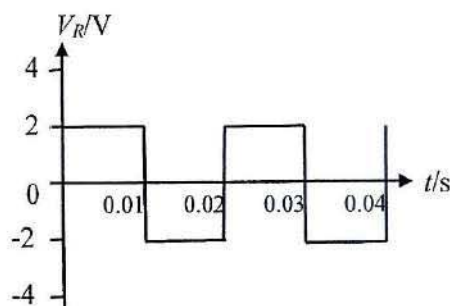
b) $V = IR$
 $3.5 = I(3.0)$ OR $I_{\text{rms}} = \frac{I_{\text{peak}}}{\sqrt{2}} = \frac{V_{\text{peak}}}{\frac{R}{\sqrt{2}}} = \frac{5.0}{3.0} = 1.2\text{ A}$

c) $P_{\text{avg}} = \frac{V_{\text{rms}}^2}{R} = \frac{3.5^2}{3.0} = 4.1\text{ W}$ OR $P_{\text{avg}} = \frac{P_{\text{peak}}}{2} = \frac{V_{\text{peak}}^2}{R} \times \frac{1}{2} = \frac{5.0^2}{3.0} \times \frac{1}{2} = 4.2\text{ W}$

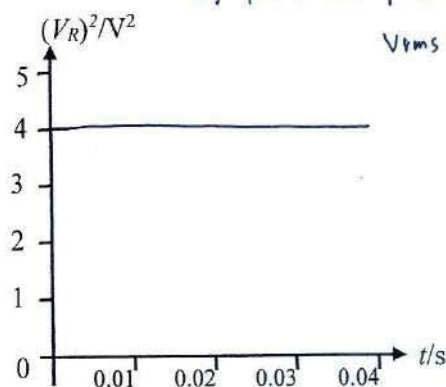
d) The rms voltage, 3.5 V

16.3.2 Root-mean-square (r.m.s.) values for non-sinusoidal AC**Example 4:**

The waveform of potential difference V_R across a $3\ \Omega$ resistor is shown below. Determine (a) the mean p.d., (b) the r.m.s. p.d. and (c) the average power dissipated in the resistor.



a) The mean p.d.: Mean $V = 0$

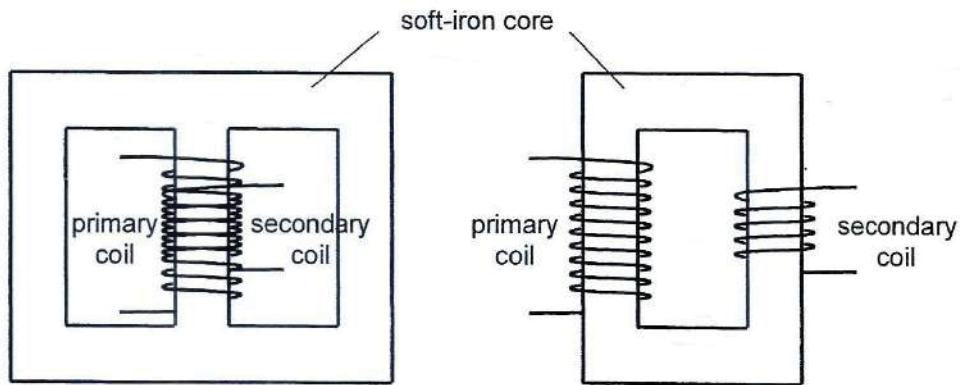


b) The r.m.s. p.d.: mean $V^2 = 4$
 $V_{\text{rms}} = \sqrt{\text{mean } V^2} = \sqrt{4} = 2\text{ V}$

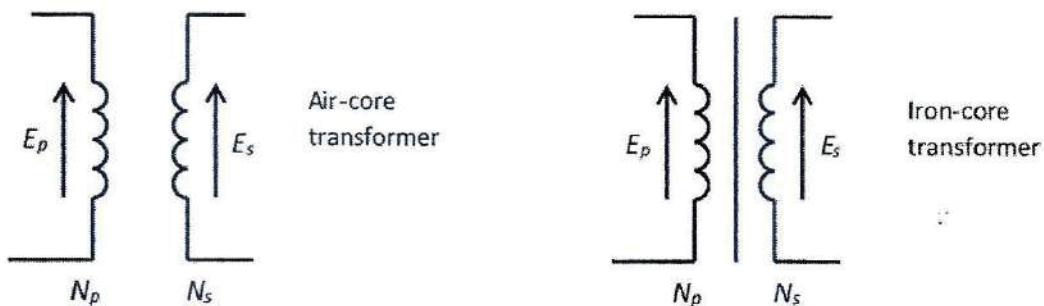
c) Avg $P = \frac{V_{\text{rms}}^2}{R} = \frac{2^2}{3} = 1.3\text{ W}$

16.4 Transformer

The transformer is a device which makes use of electromagnetic induction to produce a large alternating e.m.f. from a small one, or a small alternating e.m.f. from a large one. The primary and secondary coils are formed by winding insulated wires on a soft iron core, either one on top of the other or on separate limbs (as shown in the figures below). There is no electrical connection between the coils but they are linked magnetically via the soft iron core; i.e. the flux is the same in both coils.

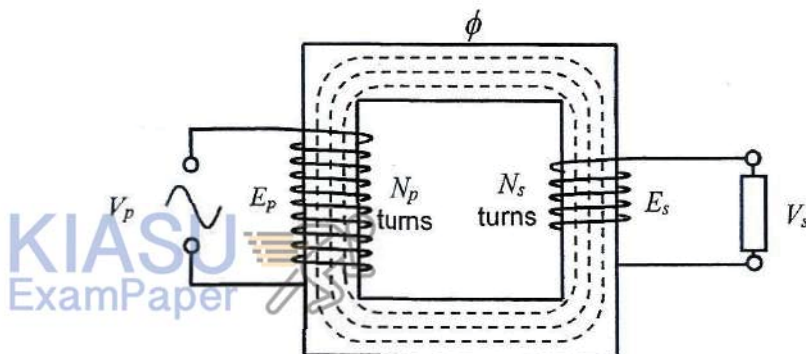


Below are circuit symbols of transformers.



16.4.1 Induced e.m.f.s in a transformer

When an alternating e.m.f. V_p is applied to the primary coil, alternating currents in both primary coil and secondary coil flow and set up a resultant alternating flux ϕ in the core.



If N_p and N_s are the number of turns in the primary and secondary coils, their linkages with the flux ϕ are

$$(\Phi_{\text{linkage}})_p = N_p \phi \quad \text{and} \quad (\Phi_{\text{linkage}})_s = N_s \phi$$

The alternating flux induces an alternating e.m.f. E_p in the primary coil as well as an alternating e.m.f. E_s in the secondary coil. By Faraday's law, the magnitudes of E_p and E_s are given by

$$E_p = \frac{d(\Phi_{\text{linkage}})_p}{dt} = N_p \frac{d\phi}{dt} \quad \text{and} \quad E_s = \frac{d(\Phi_{\text{linkage}})_s}{dt} = N_s \frac{d\phi}{dt}$$

Dividing the above two equations, we have

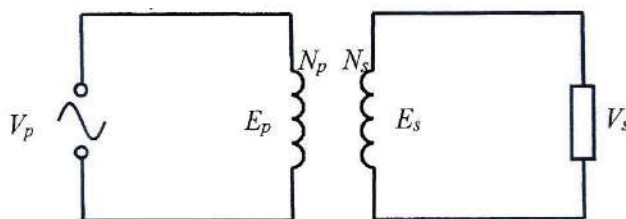
$$\frac{\text{e.m.f. induced in secondary}}{\text{e.m.f. induced in primary}} = \frac{E_s}{E_p} = \frac{N_s}{N_p}$$

Thus the transformer steps voltage up or down according to its 'turns-ratio' $\frac{N_s}{N_p}$.

If V_s is the output p.d. across the secondary coil, then for an **ideal transformer** that has zero-resistance coils,

$V_p = E_p$ and $V_s = E_s$, giving

$$\boxed{\frac{V_s}{V_p} = \frac{N_s}{N_p}}$$



Note that

- When the secondary coil is open-circuited, the same analysis can still be applied and resultant alternating flux ϕ in the core will be due to the current in the primary coil.
- V_p must be an alternating e.m.f. . If not, the flux in the primary coil would not change (except for a short time immediately after the e.m.f. is first applied) and there would be no induced e.m.f. in the secondary coil.
- The induced e.m.f. E_s has the same frequency as the applied voltage V_p .
- If $N_s > N_p$, the transformer is called a *step-up transformer* because $V_s > V_p$.
A *step-down transformer* has $N_s < N_p$.

16.4.2 Power transmission in a transformer

If the transformer is ideal, i.e. 100% efficient,

$$\text{Power output} = \text{Power input}$$

$$V_s I_s = V_p I_p$$

Thus, turns ratio

$$\frac{N_s}{N_p} = \frac{V_s}{V_p} = \frac{I_p}{I_s}$$



The above relationship is only applicable to an ideal transformer. In the A levels, transformers are assumed to be ideal unless otherwise stated.

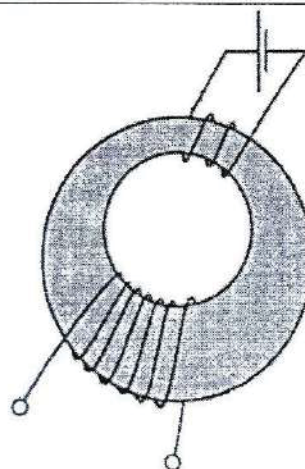
Note that the relationship is obtained from the Principle of Conservation of Energy.

Example 5:

A soft iron ring of variable cross-section has two coils wound round it. One coil has 7 turns. The other coil which has 3 turns, is connected to a d.c. supply.

Which quantity is the same inside each coil?

- ☐ A magnetic field
- ☒ B magnetic flux =
- ☐ C magnetic flux density
- ☐ D magnetic flux linkage



Example 6:

A step-up transformer has 200 turns on its primary coil and 3000 turns on its secondary coil. The primary coil is supplied with a 30 A alternating current at 90 V.

- (a) What is the voltage in the secondary circuit?
- (b) What current flows in the secondary circuit?
- (c) What is the power in the primary circuit?
- (d) What is the power in the secondary circuit?

$$a) \frac{V_s}{V_p} = \frac{N_s}{N_p} \Rightarrow \frac{V_s}{90} = \frac{3000}{200} \Rightarrow V_s = 1350 \text{ V}$$

$$b) \frac{I_s}{I_p} = \frac{N_p}{N_s} \Rightarrow \frac{I_s}{30} = \frac{200}{3000} \Rightarrow I_s = 2.0 \text{ A}$$

$$c) \langle P \rangle = I_p V_p = 30 \times 90 = 2700 \text{ W}$$

$$d) \langle P \rangle = \text{same as pri circuit assuming ideal transformer (100\% efficient)} \\ = 2700 \text{ W}$$

16.4.3 Efficiency

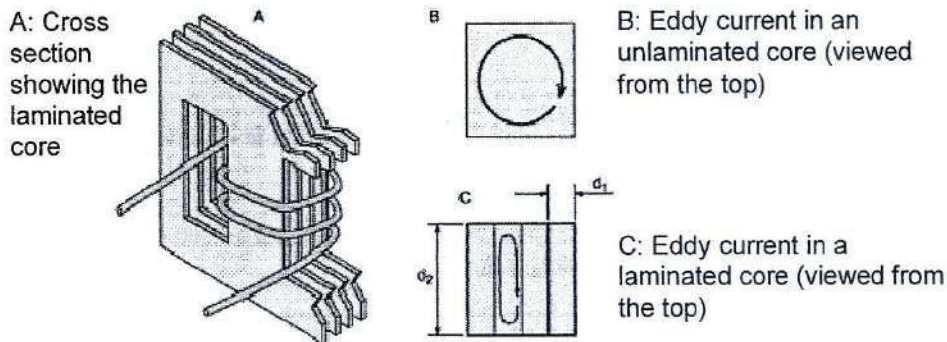
The efficiency of commercial transformers are very high – typically in the range of 95% to 99%. The four most important sources of power loss are listed below.

(a) Resistance of windings

Some energy is dissipated as heat in both coils (I^2R). This can be reduced to an acceptable level by using suitably thick wires. (The coil which has the smaller number of turns, carries the larger current. And thus, it usually requires thicker wires.)

(b) Eddy Currents

The alternating magnetic flux induces eddy currents in the iron core and causes heating. The effect is reduced by having a laminated core.



(c) Hysteresis

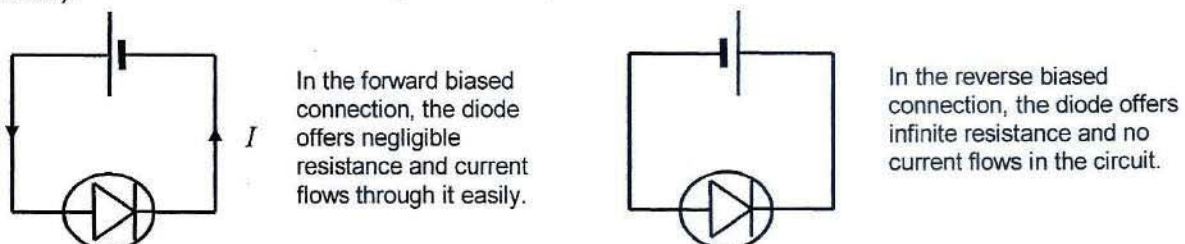
Each time the direction of magnetization of the core is reversed, some energy is wasted in overcoming internal friction. This is known as hysteresis loss and it produces heating in the core. It is minimized by using special alloys (e.g. Permalloy which is an alloy of nickel and iron) for the core material.

(d) Flux Leakage

The flux due to the primary coil may not all link to the secondary coil if the core is badly designed.

16.5 Rectification

The I - V characteristic of rectifiers (or diodes) are such that they present a very low resistance to current flow in one direction (when forward biased), and a very high resistance in the opposite direction (when reverse biased).

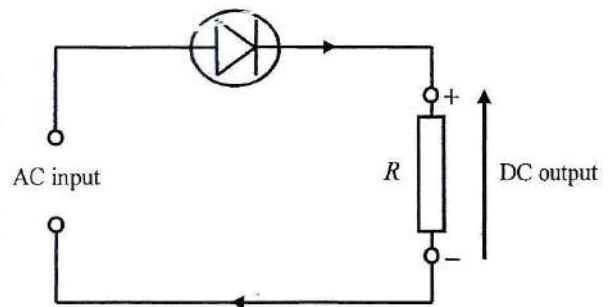


16.5.1 Half-wave rectification

Rectification is the conversion of AC to DC. A simple half-wave rectifying circuit is shown below. For simplicity, the resistor R is used to represent a load which requires a DC power source.

As the applied potential difference to the rectifier and load is a sinusoidal wave, the diode is forward biased during one half-cycle, and reverse biased during the next half-cycle.

During the half-cycle when the rectifier is forward biased, current flows round the circuit, creating a potential difference across R . This p.d. will have almost the same value as the applied p.d. if the forward resistance of the diode is small compared with R .

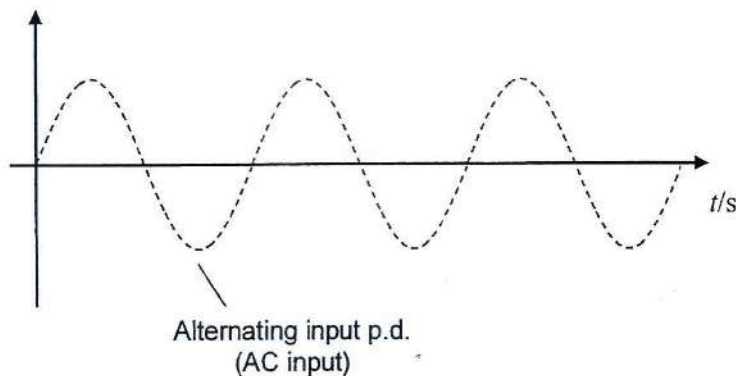


During the half-cycle when the diode is reverse biased, little or no current flows and the potential difference across R is zero.

This is repeated for each cycle of AC input. The current flow is unidirectional and so is the potential difference across R . Although it fluctuates, it never changes direction.

Example 7:

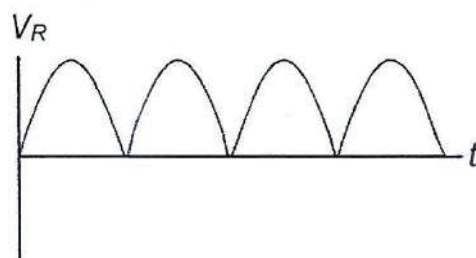
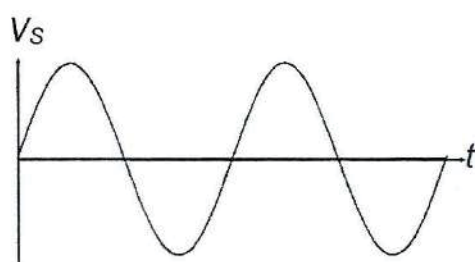
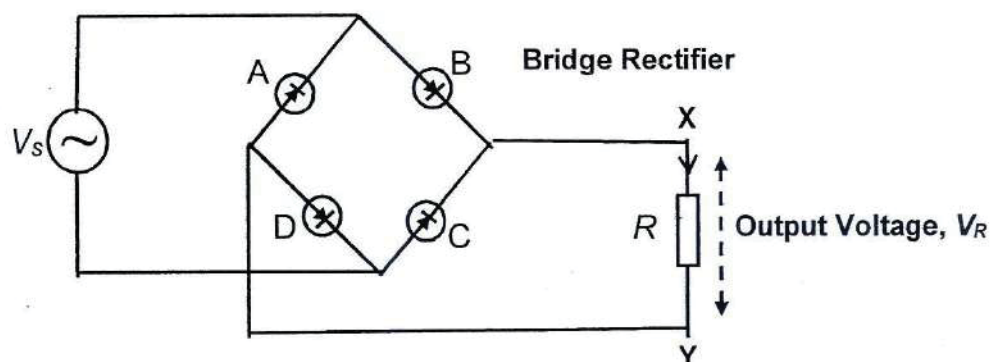
Sketch the waveforms for the current and potential difference across the load R after half-wave rectification.



Appendix

1 Full-wave Rectification

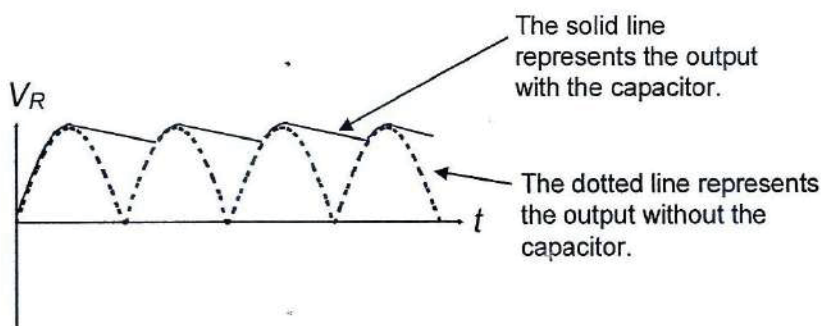
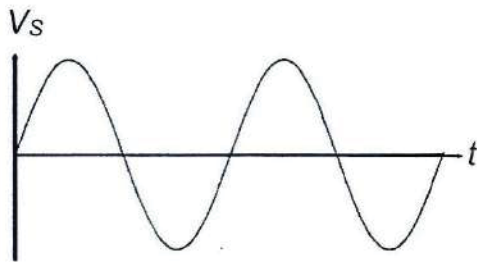
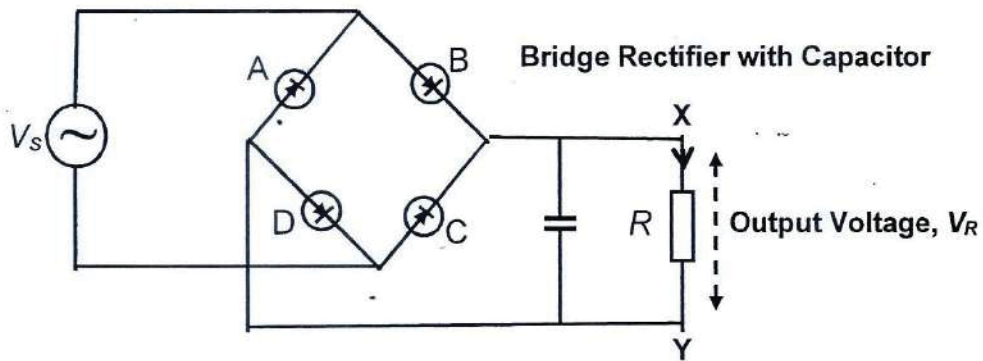
Full-wave rectification can be accomplished with a bridge rectifier consisting of 4 diodes. During any half cycle, two of the four diodes are forward biased.



- When the upper terminal of V_s is positive with respect to the lower terminal, diodes B and D are forward biased, while diodes A and C are reverse biased.
- When the lower terminal of V_s is positive with respect to the upper terminal, diodes A and C are forward biased, while diodes B and D are reverse biased.
- Hence, current will always flow in the direction from X to Y through the load resistor R , and the output voltage V_R is unidirectional.

2 Smoothing output voltage

Smoothing is the process of reducing the fluctuations of the rectified DC output. In other words, it is to make the output more constant. This is done by connecting a special device, known as a capacitor, in **parallel** with the load R .



The output of the rectifier is 'smoothed' by connecting a capacitor across the output.



Topic 17: Quantum Physics

C2 2017

Contents

Quantum Physics Syllabus	3
17.1 The Photoelectric Effect	
17.1.1 Background	4
17.1.2 Quantum Theory of Light	4
17.1.3 Experimental Setups to Investigate the Photoelectric Effect	7
17.1.4 The Photoelectric Effect	11
17.2 Atomic Structure and Line Spectra	
17.2.1 Atomic Models	18
17.2.2 Emission Line Spectra	21
17.2.3 Absorption Line Spectra	23
17.3 X-ray Spectra	
17.3.1 X-ray Production	25
17.3.2 Explanation of X-ray Emission Spectra	26
17.4 Wave-Particle Duality and The Uncertainty Principle	
17.4.1 Introduction	29
17.4.2 de Broglie's Hypothesis – Matter Waves	29
17.4.3 Evidence of Wave-like Nature of Particles	30
17.4.4 The Heisenberg Uncertainty Principle	32
Appendix Simple Analysis of Electron in a Box	34

Syllabus 9749 (H2 Physics)**Content**

- Energy of a photon
- The photoelectric effect
- Wave-particle duality
- Energy levels in atoms
- Line spectra
- X-ray spectra
- The uncertainty principle

Learning Outcomes

Candidates should be able to:

- (a) show an appreciation of the particulate nature of electromagnetic radiation.
- (b) recall and use $E = hf$.
- (c) show an understanding that the photoelectric effect provides evidence for the particulate nature of electromagnetic radiation while phenomena such as interference and diffraction provide evidence for the wave nature.
- (d) recall the significance of threshold frequency.
- (e) recall and use the equation $\frac{1}{2}mv_{\max}^2 = eV_s$, where V_s is the stopping potential.
- (f) explain photoelectric phenomena in terms of photon energy and work function energy.
- (g) explain why the stopping potential is independent of intensity whereas the photoelectric current is proportional to intensity at constant frequency.
- (h) recall, use and explain the significance of $hf = \phi + \frac{1}{2}mv_{\max}^2$.
- (i) describe and interpret qualitatively the evidence provided by electron diffraction for the wave nature of particles.
- (j) recall and use the relation for the de Broglie wavelength $\lambda = h/p$.
- (k) show an understanding of the existence of discrete electron energy levels in isolated atoms (e.g. atomic hydrogen) and deduce how this leads to the observation of spectral lines.
- (l) distinguish between emission and absorption line spectra.
- (m) recall and solve problems using the relation $hf = E_2 - E_1$.
- (n) explain the origins of the features of a typical X-ray spectrum.
- (o) show an understanding of and apply $\Delta p \Delta x \gtrsim h$ as a form of the Heisenberg position-momentum uncertainty principle to new situations or to solve related problems.

Introduction

By the middle of the 19th Century, classical physics (Newtonian mechanics and electromagnetic theory) was well established and scientists of the day were confident that they could explain many phenomena happening in a physical world. However, within a few decades (by the end of the 19th Century), it became clear that there were some problems such as several observed phenomena continued to defy explanations by classical physics:

1. Blackbody radiation and the ultraviolet catastrophe (MOE H3)
2. Photoelectric effect (H2)
3. Compton Effect (MOE H3)
4. Bright line spectrum of gas discharge tubes (H2)

While scientific attempts were focusing on trying to explain these experiments, a new theory dubbed as “quantum theory” was developed with contributions from a number of eminent scientists.

Max Planck received his Nobel Prize in 1918, for resolving the issue of blackbody radiation and ultraviolet catastrophe by introducing the concept of discrete energy (quantum) and a crucial constant h known as Planck constant. Albert Einstein received his Nobel Prize in 1921 by extending Planck’s proposal of discrete energy for electromagnetic radiation to successfully explain the photoelectric effect. Niels Bohr was honoured with a Nobel Prize in 1922 for explaining the bright line spectrum using the quantum theory of light.

Items 2 and 4, along with some introductory quantum mechanics, are covered in this topic. For a more holistic and comprehensive understanding, you are encouraged to go online to find out more about items 1 and 3.

Without the scientists and their efforts in the development of quantum theory, our world will be devoid of modern technology such as lasers and many electronic based devices.

17.1 The Photoelectric Effect

17.1.1 Background

In 1887, Heinrich Hertz discovered the **photoelectric effect** – electrons were emitted when he shone UV light on a freshly polished zinc plate.

The **photoelectric effect** is a phenomenon in which electrons are liberated from a cool metal surface when electromagnetic radiation of sufficiently high frequency is incident upon it.

Philippe Lenard performed some very careful investigations of the effect in 1902. His results could not be explained using classical physics and the wave theory of light. It was Albert Einstein who finally explained the photoelectric effect with particulate nature of light in 1905.

17.1.2 Quantum Theory of Light

In 1900 Max Planck, attempting to explain black body radiation, suggested that, although light was a wave, these waves could gain or lose energy only in finite amounts related to their frequency. Planck called these lumps of light energy *quanta*. We now call them **photons**. In 1905, Albert Einstein used the idea of light quanta to explain the photoelectric effect and suggested that these light quanta had a real existence.



An introduction to Quantum Theory of Light can be found in the following YouTube video:

A **photon** is a quantum of electromagnetic radiation. Its energy is given by

$$E_{\text{photon}} = hf$$

where h is Planck constant ($h = 6.63 \times 10^{-34} \text{ J s}$) and f the frequency of the light.

Einstein was building on the work of Max Planck. Planck first determined the value of the constant h when he was trying to explain the spectrum of blackbody radiation in 1900.

Since each photon travels at the speed of light c , E can also be written as

$$E = \frac{hc}{\lambda}, \text{ as } c = f\lambda$$

An alternative unit for Energy

On the atomic scale, the magnitude of energy involved is very small. Hence, we find it convenient to define an alternative unit of energy called the electron volt (eV). The electron volt is defined as the energy the charge-field system gains or loses when a charge of magnitude e is moved through a potential difference of one volt. In other words, when a particle of charge e moves through a potential difference of 1 V, the change in the potential energy is 1 eV.

$$1 \text{ eV} = 1.6 \times 10^{-19} \text{ J}$$

Example 1

The eye is most sensitive to light of wavelength 550 nm. What is the energy of photons with this wavelength? Express your answer in electron volts (eV).

$$\begin{aligned} hf &= h \frac{c}{\lambda} \\ &= (6.63 \times 10^{-34}) \left(\frac{3 \times 10^8}{550 \times 10^{-9}} \right) \\ &= 3.62 \times 10^{-19} \text{ J} \\ &= (3.62 \times 10^{-19} \text{ J}) \left(\frac{1 \text{ eV}}{1.6 \times 10^{-19} \text{ J}} \right) \\ &= 2.26 \text{ eV} \end{aligned}$$

KIASU
ExamPaper

Example 2

A source emits monochromatic light of wavelength 550 nm at a power of 1.25 W. How many photons are emitted in one second?

$$E = Nhf \quad \text{--- (1)}$$

$$E = P \times t \quad \text{--- (2)}$$

$$\begin{aligned} N &= \frac{Pt}{hf} \\ &= \frac{(1.25 \text{ W})(1 \text{ s})}{(3.62 \times 10^{-19} \text{ J})} \\ &= 3.453 \times 10^{18} \end{aligned}$$

17.1.3 Experimental Setups to Investigate the Photoelectric Effect

Let's look at some investigations of the photoelectric effect before using the "photon" theory to explain it. Fig. 1 shows a schematic diagram of a typical set-up used to investigate the photoelectric effect.

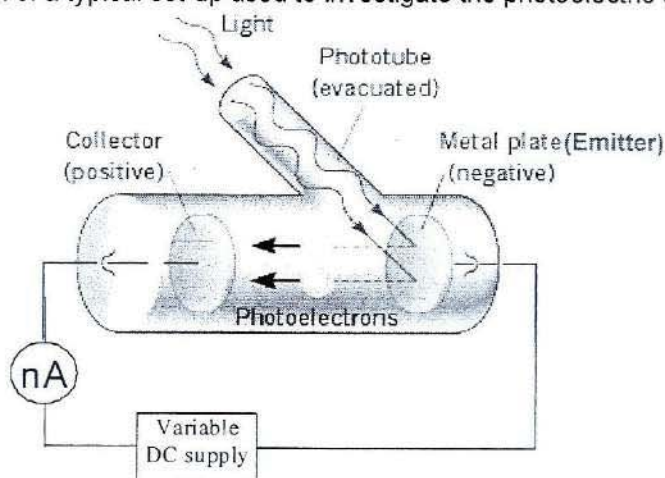


Fig. 1 Experimental set-up

Two metal plates are located at two ends of an evacuated tube. The two metal plates can be kept at different potential differences and polarity by an adjustable DC supply.

Light of various **frequencies** and **intensities** are shone on one of the metal plates (the emitter). In doing so, electrons **could** be ejected from the emitter and travel to the other metal plate (the collector).

If electrons reached the collector, a current would be registered and measured by the ammeter as shown in the diagram.

The ejected electrons are also called **photoelectrons**.

Experiment 1: Varying the Applied Voltage

A monochromatic (single wavelength) source of EM radiation of constant intensity is shone on the emitter.

When the adjustable DC supply is adjusted such that the collector is at a **higher** potential than the emitter, an electric field is set up to attract the photoelectrons towards the collector (Fig. 2(a)), and a current can be detected.

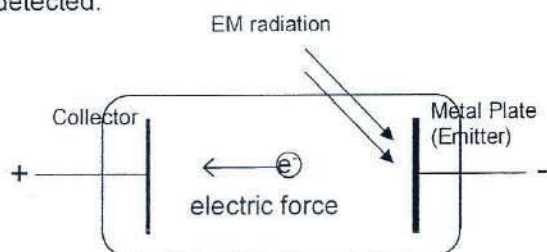


Fig. 2(a)

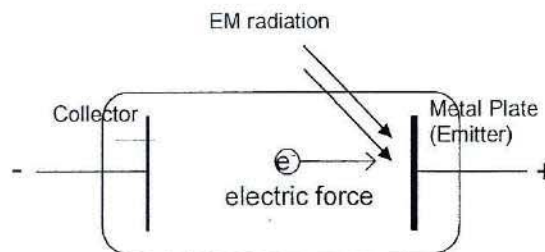


Fig. 2(b)

When the adjustable DC supply is adjusted such that the collector is at a **lower** potential than the emitter, it will produce an electric field that repels the photoelectrons away from the collector (Fig. 2(b)). A small potential difference is not able to stop the more energetic photoelectrons from reaching the collector. Hence, a current can still be detected by the ammeter if the potential difference is not large enough.

By adjusting the DC supply, there will be a value such that the current becomes zero – this corresponds to the potential difference whereby even the **most energetic** photoelectrons could not reach the collector. This potential difference across the adjustable DC supply is known as the **stopping potential**.

Results and Interpretation

Typical results are shown in Fig. 3.

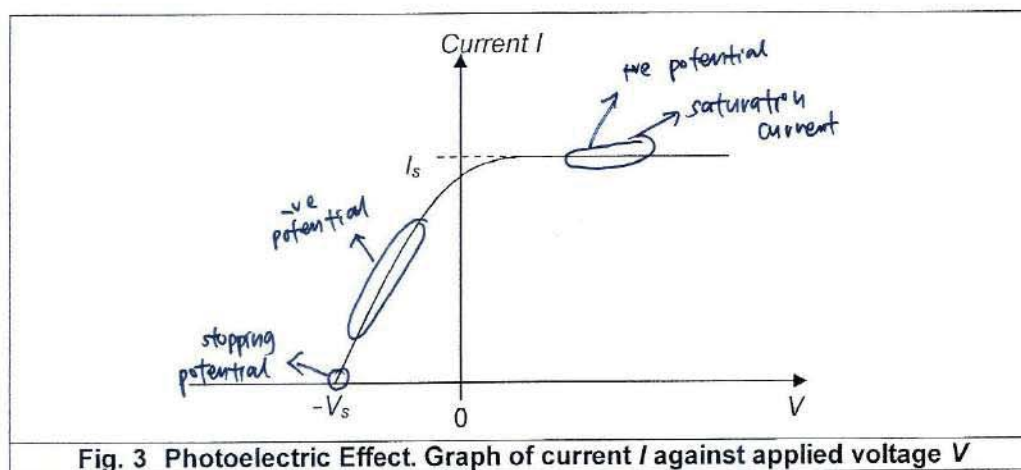


Fig. 3 Photoelectric Effect. Graph of current I against applied voltage V

Saturation Current, I_s

The applied voltage V is the potential of the collector with respect to the emitter. As V is increased, the current I increases as well. It reaches a maximum and remains constant a little beyond $V = 0$. Increasing V further does not increase I . This maximum current is known as the saturation current I_s . It represents the maximum rate of emission of photoelectrons from the emitter as all ejected photoelectrons reach the collector.

Note: The reason that the current reaches a maximum a little beyond $V = 0$ and not exactly at $V = 0$ is that the electrons are not all aimed straight at the collector. However, in most A-Level questions, the saturation current is reached at $V = 0$. See, for instance questions S7 and D3 in Tutorial 17A.

Stopping Potential (a measure of maximum KE of the photoelectrons), V_s

When V becomes more negative, the current decreases gradually until it becomes zero at a certain potential difference. The magnitude of this potential difference is known as the stopping potential V_s .

The current decreases gradually because when the photoelectrons are ejected from the metal surface, they can have different kinetic energies (KE). When the applied voltage is negative, the photoelectrons are repelled by the collector, but some can still reach the collector if they have sufficient KE .

As the applied voltage becomes more negative, more and more photoelectrons which do not have sufficient KE will fail to reach the collector and the current decreases. When the negative applied voltage is large enough, even the most energetic photoelectrons cannot reach the collector, i.e., $I = 0$.



The adjustable DC supply is used as a means to determine the KE possessed by the most energetic photoelectrons (i.e., KE_{max}) when they are first ejected from the metal surface. The applied potential difference V does not alter the KE_{max} in any way.

The potential difference just enough to stop even the most energetic photoelectrons is called the **stopping potential V_s** . By applying the Principle of Conservation of Energy, the kinetic energy of the most energetic photoelectrons KE_{max} , is therefore equal to eV_s .

Thus the stopping potential is a measure of the maximum kinetic energy (KE_{max}) of the photoelectrons.

$$KE_{max} = eV_s$$

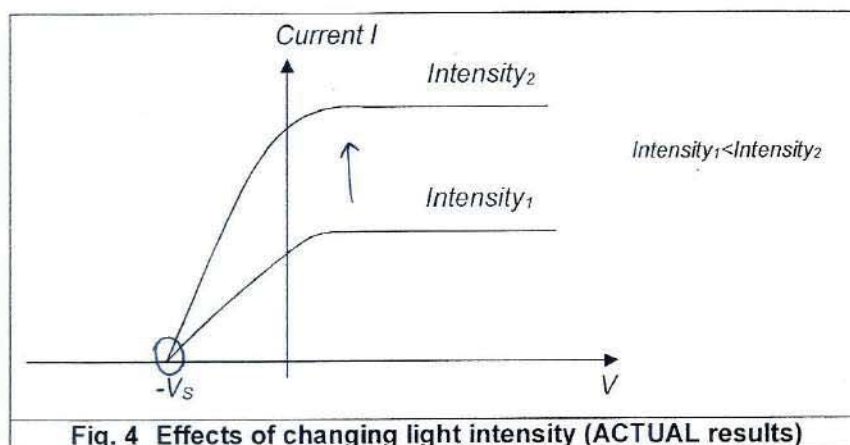
Now that we have a way of measuring KE_{max} of the photoelectrons, we can investigate how KE_{max} changes with other factors like the intensity and frequency of the incident light.

Experiment 2: Varying the Applied Voltage for Different Intensities of Light

Experiment 1 is repeated for two different intensities of the incident light, $Intensity_1$ and $Intensity_2$. The frequency of the incident light remains the same.

Results and Interpretation

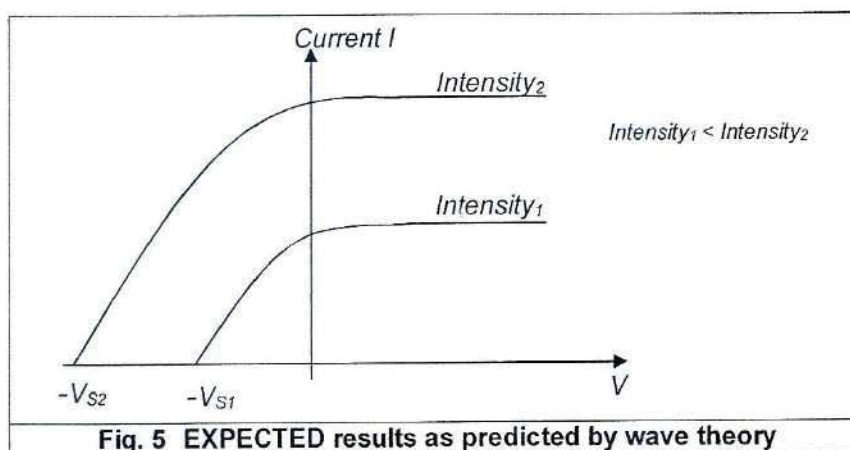
Typical results are shown in Fig. 4.



The effect on the saturation current (I_s) is higher for increased intensity, $Intensity_2$.

However, the graph also shows that KE_{max} is not affected by the intensity. This observation is not consistent with the wave theory of light.

According to the wave theory, increasing intensity would increase **both** I_s (saturation current) and KE_{max} (hence the stopping potential) of the photoelectrons. The wave theory would predict results similar to Fig. 5.



17.1.4 The Photoelectric Effect

Einstein offered an explanation for the photoelectric effect by assuming light to be made up of photons. According to Einstein, when light shines on a metal, a photon can interact with an electron and causes the electron to be ejected from the metal.

In the interaction, the photon gives up all its energy to the electron and disappears in the process. Part of this energy is used to do work necessary for the electron to leave the metal. The rest of the energy appears as the kinetic energy of the ejected electron. Each ejected electron is the result of an interaction with **one** photon only.

Some electrons are more strongly held than others. For the **least** strongly held electrons, the work done is a minimum and is known as the **work function** of the metal, Φ . These least strongly held electrons will have maximum KE when released through interaction with a photon.

The **work function** is the minimum amount of energy required to remove an electron from the metal surface (i.e. the energy required to remove the least strongly held electron).

Different metals have different lattice structures and hence different work functions.

By applying the Principle of Conservation of Energy to a single photon-electron interaction:

Energy of photon absorbed = Work done to liberate an electron from material + KE of emitted electron

For the **least strongly held** electron:

hf	=	Φ	+	KE_{max}
photon energy		energy needed to eject the least strongly held electron		kinetic energy of the least strongly held electron when ejected

Einstein's proposal clearly predicts that the KE_{max} is dependent on the frequency but not the intensity of the incident radiation.

It also clearly explains why a **threshold frequency** exists. The threshold frequency is the minimum frequency of the incident radiation required to cause the emission of electrons from the metal.

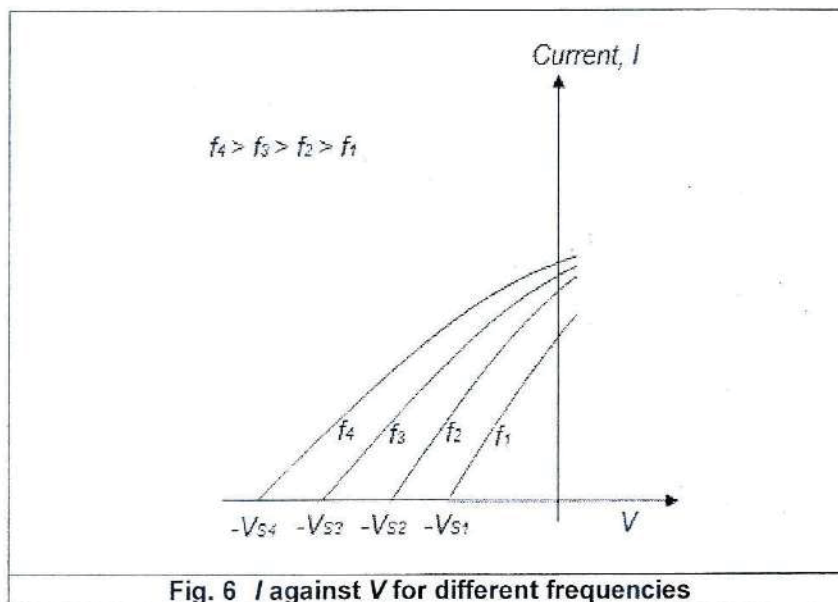
If the energy of a photon is lower than the work function (i.e. $hf < \Phi$), then the photon can never eject an electron from the metal. Therefore, below a certain threshold frequency, no electrons can be emitted.

Normally one photoelectron is ejected by interacting with one photon. The probability of two photons simultaneously hitting a single electron is very low.

Experiment 3: Varying the Applied Voltage for Different Frequencies of Light

In this experiment, we will repeat Experiment 1 for different frequencies. For a given value of f , we vary the applied potential difference, V and measure the current, I . By plotting I against V , the stopping potential, V_s , can be determined (i.e. the value of V when $I = 0$).

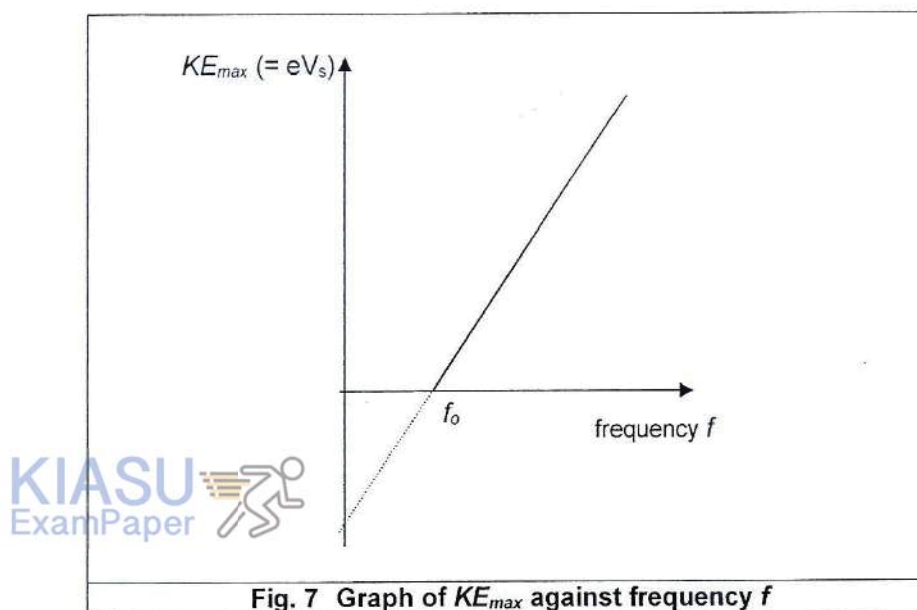
We will repeat this process for different values of incoming frequencies.



However, KE_{max} cannot be measured directly in this experiment. If we want to plot KE_{max} against f , we use the relationship between KE_{max} and the stopping potential on page 9, which is given by:

$$KE_{max} = eV_s$$

We may now plot each value of f against its corresponding value of eV_s ,



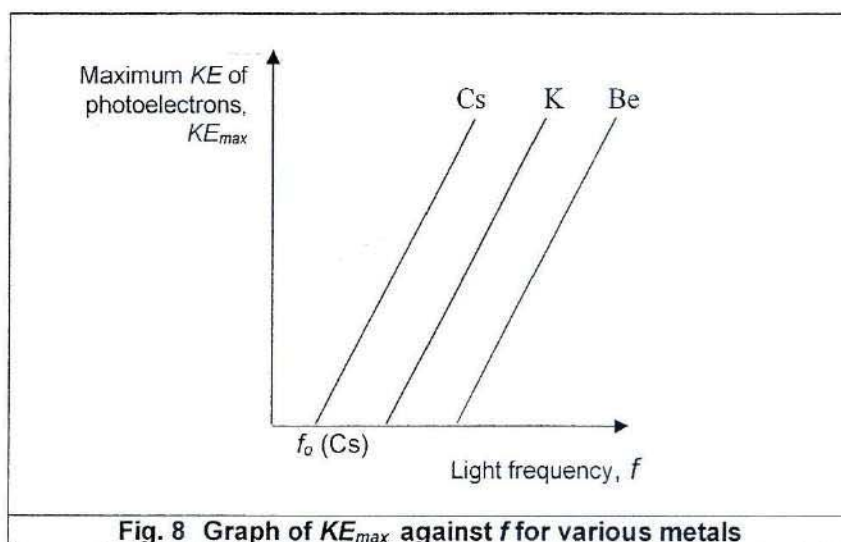
Validation of Relationship

The results indicate that the relationship

$$KE_{\max} = hf - \Phi$$

is correct. This gives strong support for Einstein's theory for the photoelectric effect.

When the experiment is repeated for different metals (i.e, different Φ), Fig. 8 is obtained.



The gradient is the same for all the metals, but the x and y intercepts are different.

Drawing Conclusions

From the equation:

$$KE_{\max} = hf - \Phi$$

(i) The gradient and the Planck constant

The gradient gives the value of Planck constant.

The gradient is the same for different metals and should be equal to the Planck constant, a universal constant.

Experimental data shows that the gradient of the graph agrees with the value for h which Planck had derived separately when he was trying to fit experimental data for the spectrum of black body radiation.

This is very strong evidence in support of Einstein's theory because experimental data from different metals gave the same value for h .

(ii) The vertical-intercept and the work function

The value of the vertical-intercept gives the work function of the metal.

From the equation, the y intercept gives the work function of the metal. The value varies for different metals as they have different lattice structures. The work function is usually given as a positive number. So, for example, if y intercept is -2.80 eV , $\Phi = 2.80 \text{ eV}$.

(iii) The x intercept and the threshold frequency

When $KE_{\max} = 0$, the energy hf_0 of the incident photon is equal to the work function Φ of the metal:

$$hf_0 = \Phi$$

This means that according to the "photon" theory, if the incident light has a frequency below f_0 , no photoelectrons will be produced, no matter how intense the light.

$f_0 \text{ is the threshold frequency, } f_0 = \Phi / h$

For different metals, the threshold frequency is different because different metals have different work functions.

Summary - How the "Photon" Theory explains the observations of the Photoelectric Effect

The three problems below show the failure of the wave theory and the success of the photon (or particle) theory of light in explaining the observations from photoelectric emission.

1. The Intensity Problem

Wave Theory: According to the wave theory of light, when the intensity increases, there is a larger amount of energy in the wave and the kinetic energy of the photoelectrons would be expected to be higher. This would result in a higher measured KE_{max} . Experiments show that this is not true; intensity does not affect KE_{max} .

Photon Model: If intensity is increased, keeping frequency constant, the number of photons falling on unit area per unit time is increased but the energy of each photon remains the same. Since each photoelectron is produced by one photon, the maximum kinetic energy of the electron remains unchanged.

Recall the definition of intensity,

Intensity = power passing per unit area perpendicular to direction of energy propagation

$$\text{Intensity} = \frac{\text{power}}{\text{Area}} = \frac{\text{Energy}}{\text{time} \times \text{area}} = \frac{N \times (hf)}{t \times A}$$

where N : No. of photons; hf : energy of 1 photon; t : time; A : cross sectional area

If the frequency remains constant, then energy of each photon remains the same. **Greater intensity simply means larger number of photons incident per unit time.** This just means that a proportionately larger number of electrons will be emitted per unit time, resulting in a proportionately larger photoelectric current. **The KE_{max} of the photoelectrons does not change.**

2. The Frequency Problem

Wave Theory: The wave theory cannot explain the existence of the threshold frequency. According to the wave theory, if the intensity is sufficiently high, electrons will be ejected regardless of the frequency. There should not be a threshold frequency.

Photon Model: If the energy of the photons is below hf_0 , the photons cannot liberate any electrons. The minimum amount of energy needed to liberate an electron is Φ , which is equal to hf_0 , where f_0 is threshold frequency.

3. The Time Delay Problem

Wave Theory: The wave theory predicts that at very low intensities, there will be a time delay before photoelectrons are detected. This is because at the low rate of transfer of energy to the electron, ample time needs to be given for sufficient energy to be accumulated for the electron to escape from the metal. However, experiments show that there is negligible time delay. Photoelectrons are detected as soon as the metal is exposed to light of suitable frequency, even if the intensity is low.

Photon Model: The energy of a photon is transferred in a single interaction. If the photon energy is large enough, electrons should be liberated as soon as the light shines on the metal. There should be no time delay.

Example 3

In photoemission from a given material, when the wavelength of the incident light is increased, the

- A maximum kinetic energy of the emitted electrons increases.
- B minimum kinetic energy of the emitted electrons decreases.
- C minimum kinetic energy of the emitted electrons increases.
- ☒ D maximum kinetic energy of the emitted electrons decreases.

Example 4 (N89/II/29)

In an experiment on the photoelectric effect, different metals were illuminated with monochromatic light of variable frequency and variable intensity. It was found that, for each metal, electrons were emitted with a spread of kinetic energies up to a maximum value. This maximum kinetic energy depends on

- A the metal but not on the frequency or the intensity.
- B the metal and the intensity but not on the frequency.
- ☒ C the metal and the frequency but not on the intensity.
- D the frequency and the intensity but not on the metal.
- E the metal, the frequency and the intensity.

Example 5 (N84/II/14)

A student connects a photocell to a supply and finds that when the cell is exposed to monochromatic radiation a current flows only when the potential difference across the cell is less than 1.6 V. What is the maximum energy of the emitted electrons?

When there is a p.d. across the cell, the KEs of the photoelectrons are converted PEs as they approach the collector. 0 current is observed if the most energetic photoelectrons are not able to reach the collector.

Based on Einstein's photoelectric equation, the most energetic emitted electrons has an energy given by the diff b/w the energy of the photons & the work function of the material.

$$KE_{\max} = eV_s = 1.6 \text{ eV}$$

Example 6 (N90/II/29)

Light quanta of energy $3.5 \times 10^{-19} \text{ J}$ fall on the cathode of a photocell. The current through the cell is just reduced to zero by applying a stopping potential of 0.25 V. What is the work function energy of the cathode?

A photon is a quantum of light energy.
Light quanta refers to a stream of photons

Based on Einstein's Photoelectric eqn.

$$\begin{aligned} \text{Work function, } \phi &= hf - eV_s \\ &= 3.5 \times 10^{-19} - (1.6 \times 10^{-19})(0.25) \\ &= 3.1 \times 10^{-19} \text{ J} \end{aligned}$$

Example 7 (N91/I/28)

An ultra-violet source of radiation causes the emission of electrons from a zinc plate. What would be the effect of a more intense source of the same wavelength on the maximum energy per electron, and on the number of electrons emitted per second?

	<u>Max. energy per electron</u>	<u>No. of electrons emitted per second</u>
A	less	more
B	the same	the same
☒ C	the same	more
D	more	the same
E	more	more

17.2 Atomic Structure and Line Spectra

17.2.1 Atomic Models¹

For a long time, people have been wondering what matter was made of. The first atomic theory of matter was developed as early as 500 BC by Democritus, in which he believed that matter was made up of particles so small that they could not be further subdivided. These particles are known as atoms (*ἄτομον*, meaning "indivisible" in Greek).

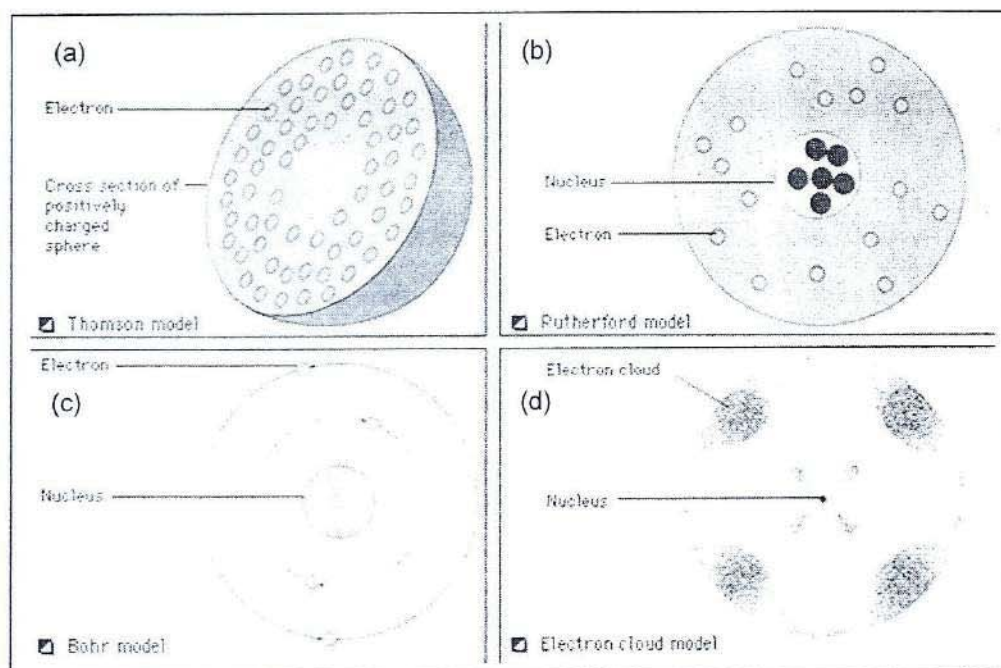


Fig. 9

Joseph John Thomson, after he discovered the existence of electrons, came up with the plum pudding model of the atom. He suggested an electrically neutral atom as composing of electrons surrounded by a relatively large, uniformly distributed, positive mass to balance the electron's charge, much like raisins embedded in a plum pudding (a traditional English Christmas fare).

The Plum Pudding model (Fig. 9(a)), however, was disproved by an experiment known as the gold foil experiment (Geiger and Marsden, 1909) whereby alpha particles (helium nuclei) are scattered at different angles when they are incident upon a thin gold foil. Ernest Rutherford proposed a model in which the electric charge of an atom is not distributed over the whole area of the atom but is concentrated in a small area (Fig. 9(b)). He envisaged the atom as having a central positive nucleus surrounded by negative orbiting electrons, somewhat similar to the solar system. This model suggested that most of the mass of the atom was contained in the small nucleus, and that the rest of the atom was mostly empty space. He tested his model successfully against data from the gold foil experiment and came up with the Rutherford scattering formula.

Unfortunately, Rutherford's model faced a very fundamental problem. Maxwell's electromagnetic theory predicted that a charge undergoing acceleration will radiate EM waves, losing energy in the process. This means that the orbiting electrons, which undergo centripetal acceleration, will lose energy through EM radiation and rapidly spiral into the nucleus!

Bohr's model of the atom

¹ In this topic, the Bohr model of the atom will be the main model used. Alternate models and the gold foil experiment will be covered in greater detail in the topic of Nuclear Physics.

In 1913, Niels Bohr presented a new model of the atom to address this problem. Bohr adopted Rutherford's picture of an atom as a nucleus surrounded by electrons moving in circular orbits, but he improved upon the old model by making a number of assumptions in order to combine the new quantum ideas of Planck and Einstein with the traditional description of a particle in uniform circular motion (Fig. 9(c)).

Inspired by the quantisation of energy levels of a blackbody in Planck hypothesis, Bohr assumed that the electrons could only orbit the nucleus in particular circular orbits with fixed energy (and angular momentum). Electrons in quantised orbits do not radiate electromagnetic waves because they could not lose energy in a continuous manner, and hence they were not allowed to spiral into the nucleus.

To incorporate Einstein's photon concept, Bohr theorized that a photon is emitted only when an electron jumps orbit from a larger one with a higher energy to a smaller one with a lower energy (Fig. 10).

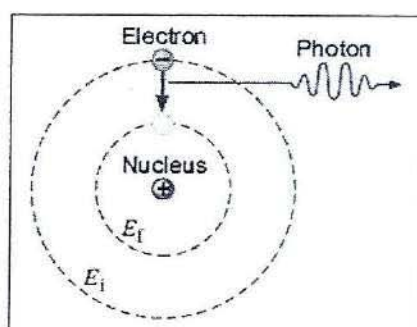


Fig 10

When an electron in an initial orbit with a larger energy E_i changes to a final orbit with a smaller energy E_f , the emitted photon has an energy of $E_i - E_f$, consistent with the law of conservation of energy.

$$hf = E_i - E_f$$

where f is the frequency of the emitted light. This process is known as **de-excitation**.

Electrons in the atom can also jump from a lower-energy orbit to a higher-energy orbit by acquiring energy (this is known as **excitation**):

- through collision of the atom with a colliding particle (such as another electron produced in a high voltage discharge tube). The colliding particle will impart energy to the atom equal to the difference between two energy levels in the atom; any remaining energy will be retained by the colliding particle as its kinetic energy.
- through absorbing a photon of energy **exactly** equal to the energy difference between two levels of the atom.

The Bohr model is sometimes known as the semi-classical model of the atom, as it adds some primitive quantization conditions to what is otherwise a classical mechanics treatment. The Bohr model is certainly not a full quantum mechanical description of the atom. The full quantum mechanical model, or electron cloud model, describes the electron as a three-dimensional shape ("cloud") where there is at least a minimal probability of finding the electron (Fig. 9(d)).

The Hydrogen atom – an example of the Bohr model



You may find a YouTube video about the Atomic Line Spectrum of Hydrogen here:

The simplest Bohr model would be that of the hydrogen atom, which consists of an electron and a proton. Bohr was able to derive the energy levels E_n for the hydrogen atom²:

$$E_n = -\frac{13.6}{n^2} \text{ eV}$$

where n is known as the principal quantum number and is a positive integer (i.e., 1, 2, 3, ...).

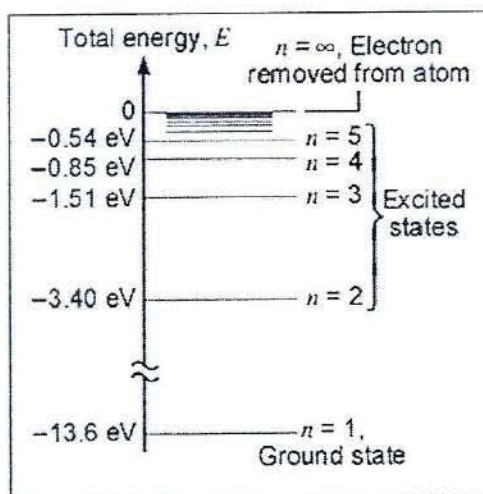


Fig. 11

The electron normally occupies the energy level corresponding to $n = 1$, also known as the ground state. At the ground state, the electron has the lowest energy, -13.6 eV, and orbits nearest to the proton (hydrogen nucleus). The higher levels are called excited states. Observe how the energies of the excited states come closer and closer together as n increases.

The electron in a hydrogen atom at room temperature spends most of its time in the ground state. To raise the electron from the ground state ($n = 1$) to the highest possible excited state ($n = \infty$), 13.6 eV of energy must be supplied. Supplying this amount of energy removes the electron from the atom, producing the positive hydrogen ion H^+ . 13.6 eV is known as the ionisation energy of the hydrogen atom.

In general, for a multi-electron atom,

The **ionisation energy** is the minimum energy needed to remove the outermost electron from the atom.

Evidence that the energy states of the atoms are discrete and quantised is provided by the existence of atomic line spectra, the subject of the next two sections.

² You are not required to memorise this formula or know how Bohr arrived at it. In fact, you are not required to know the energy levels or spectra for any particular element.

17.2.2 Emission Line Spectra

When gases such as hydrogen, neon or sodium are placed inside a narrow discharge tube at low pressure (so that the gas atoms are effectively isolated from one another), a high voltage (e.g. 1 kV) applied across its ends causes the gas atoms to ionise, forming ions and electrons. The ions and electrons accelerate towards opposite ends of the tube, under the applied electric field. Collisions between the electrons and atoms result in transfer of energy from the electrons to the atoms and the atoms are excited from their ground state. Excited states are not stable, however, and an excited atom will quickly de-excite to lower-energy states. The excited atom can, but may not de-excite directly to ground state; it may make a series of jumps through intermediate energy states to ground state, releasing more than one photon in the process.

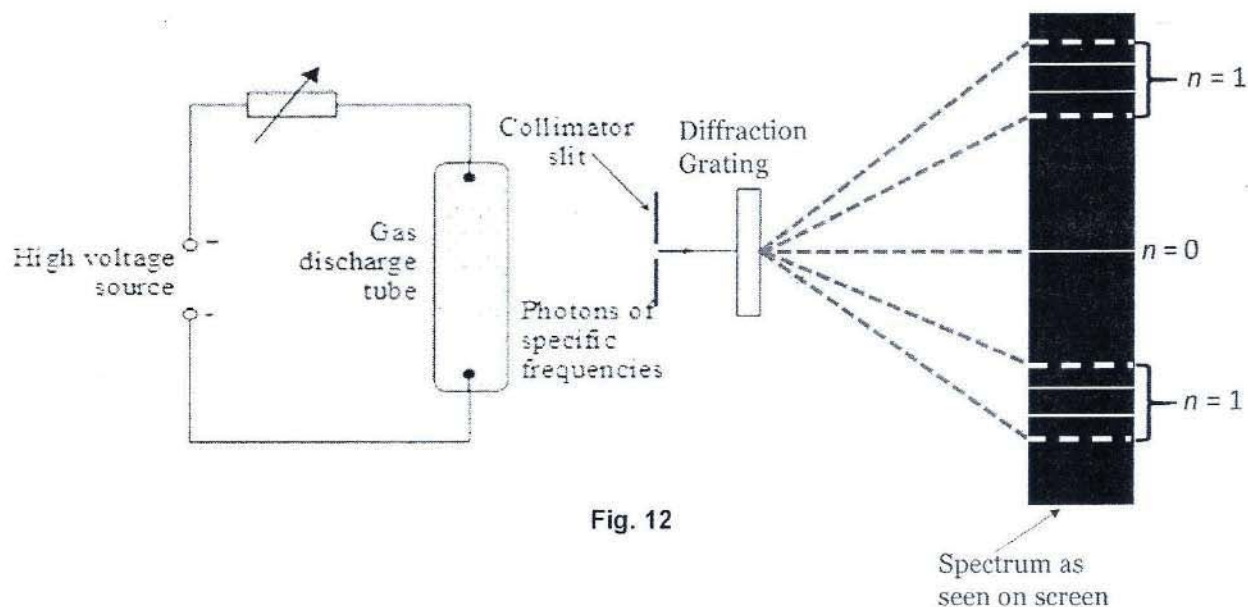


Fig. 12

The photons emitted by the gas discharge tube are passed through a collimator slit, to produce a narrow beam of photons. Next, the photons are passed through a diffraction grating to produce an **emission line spectrum** on a dark screen behind the grating.

The frequency, f , of an emitted photon depends on the energy difference between the initial and final energy states, as given by $hf = E_i - E_f$. The frequency or wavelength of the photons emitted can be determined by studying the spectrum seen on the screen (Fig. 12).

The emission line spectrum consists of a series of **distinct coloured lines on a dark background**. The different colours or frequencies are unique to the gas, i.e., no two different gases emit the same set of discrete frequencies of light and atoms of the same element always emit the same set of discrete frequencies of light.

Using his model for the energy levels of the hydrogen atom, Bohr was able to account for the spectra pattern of hydrogen gas. For hydrogen gas, several distinct groups of lines can be detected, though only one group is in the visible range (Fig. 13).

The Lyman (ultraviolet) series of lines corresponds to electron transitions down to the level $n = 1$, the Balmer (visible) series are produced by transitions down to $n = 2$, and the Paschen (infrared) series are produced by transitions down to $n = 3$; each series being named after its discoverer. The Balmer series is shown in Fig. 14.

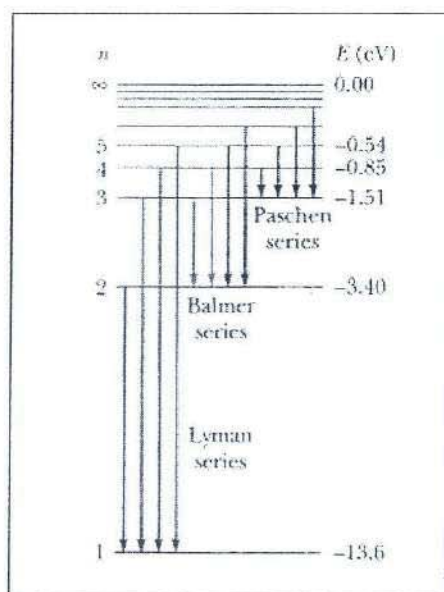


Fig. 13

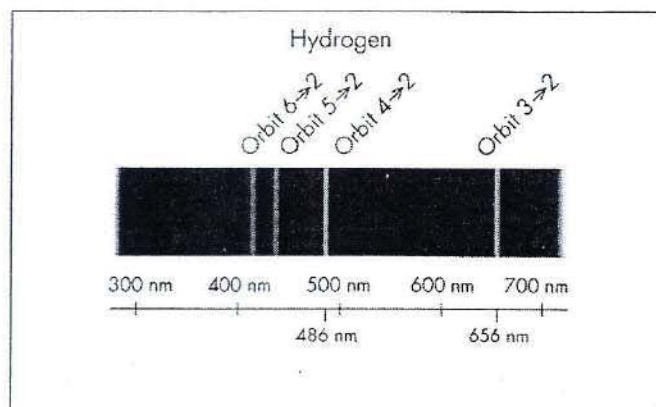


Fig. 14

When the electron is given an amount of energy equal to the ionisation energy, the electron is just able to become a 'free' electron. If the electron is given an amount of energy greater than the ionisation energy, the excess energy becomes the kinetic energy of the free electron outside the atom. **The free electron can have a continuous range of energies. Inside the atom, however, it can have only one of the energy level values characteristic of the atom.**

Typically, gas at low pressure is used for spectral analysis as the atoms are sufficiently far apart and have minimal interaction with one another. When the atoms are close together, the discrete energy levels will become several continuous energy bands due to interactions between the atoms (you may read on "Band Theory"), and the distinctive spectral lines will not be observable.

17.2.3 Absorption Line Spectra

An absorption spectrum is formed when the continuous spectrum emitted from an incandescent lamp (or white light) passes through a cool gas. The electrons in the gas absorb the frequencies corresponding to the energies they need to jump from a lower energy state to a higher energy state. These vacancies show up as "black lines" in the otherwise continuous spectrum of colours. An absorption spectrum, in a sense, is the inverse of an emission spectrum of the same series.

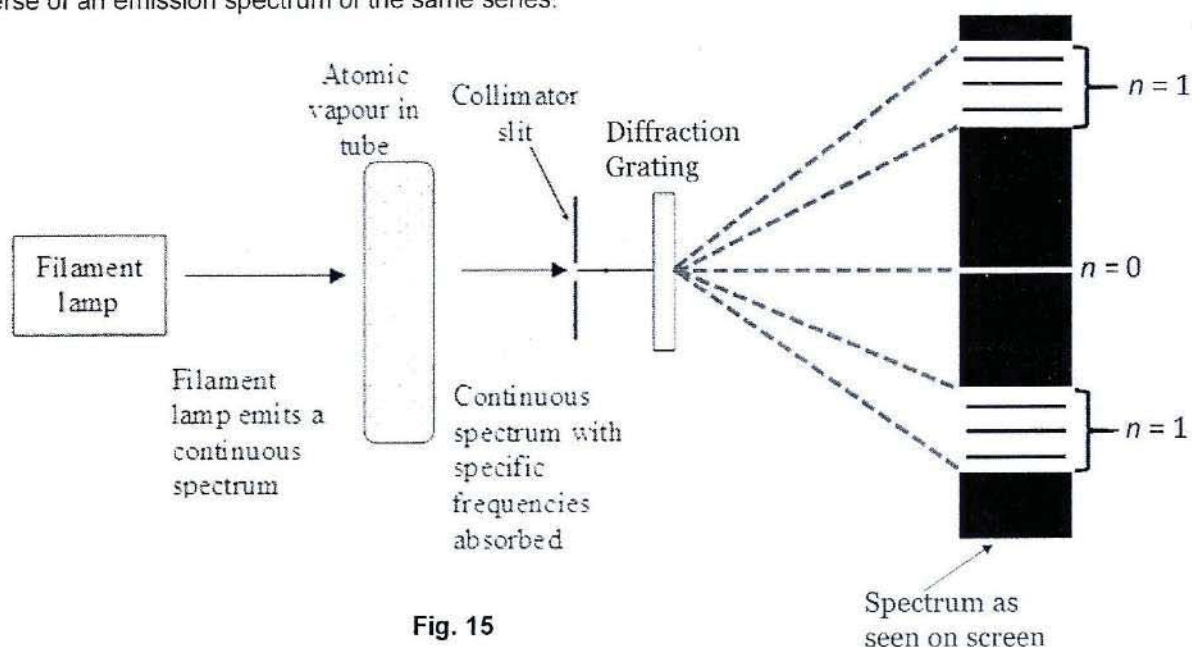


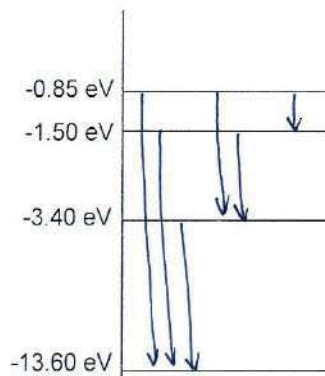
Fig. 15

In the set-up shown in Fig. 15, a filament lamp emits a continuous spectrum of light which is incident on a tube containing cool atomic vapour at low pressure - "cool" meaning atomic vapour in ground state. As a photon is incident on an atom, if its energy corresponds to the energy difference between the ground state and an excited state, it may be absorbed by the atom. As such, the atomic vapour absorbs photons of specific energies/frequencies.

What is observed on the screen is the original **continuous spectrum emitted by the hot filament lamp with darker lines corresponding to the frequencies absorbed**. Since excited states are not stable, the atoms eventually return to the ground state. However, **the photons are emitted in all directions** and not just in the forward direction. Hence the parts of the spectrum corresponding to these photon frequencies appear darker by comparison with the other frequencies.

Example 8

The diagram represents the energy levels of the four lowest states of a hydrogen atom. The energies are in units of electron volts.



- (a) Calculate the longest wavelength which might be emitted by a spectral transition between any pair of these four levels.
- (b) Calculate the excitation energy for a jump from the ground state to the next higher level. Express your answer in J.
- (c) Determine the total number of spectral lines which might be detected in the emission spectrum of atomic hydrogen, due to transitions between these four states.
- (d) Suggest why the energies of the energy states are negative.

a) Longest wavelength means smallest energy

c) 6 lines

$$E = \frac{hc}{\lambda}$$

$$\lambda = \frac{hc}{E}$$

$$= \frac{(6.63 \times 10^{-34})(3 \times 10^8)}{[-0.85 - (-1.50)] \times (1.6 \times 10^{-19})} = 1.91 \times 10^{-6} \text{ m}$$

b) $\Delta E = [-3.40 - (-13.60)] = 10.20 \text{ eV}$

$$= (10.20 \text{ eV}) \left(\frac{1.6 \times 10^{-19} \text{ J}}{1 \text{ eV}} \right) = 1.632 \times 10^{-18} \text{ J}$$

d) Remark: Same reasons as gravitational potential energies.

The reasons are:

1. Potential energy at infinity (free electron or ionized atom) is defined to be zero.

2. Potential energy is defined as work done by an external force to bring the electron from infinity to a particular location. Since the force b/w nucleus & electron is attractive, work done by external force & hence potential energy is -ve.

17.3 X-ray Spectra

17.3.1 X-ray Production

X-rays was discovered by Wilhelm K. Röntgen in 1895. He found that a gas-discharge tube which he was using with a high voltage had caused some wrapped photographic plates to become fogged. He discovered that the tube emitted a radiation that could penetrate paper, wood, glass, rubber and even thick aluminium. Röntgen could not find out whether the radiation was a stream of particles or a train of waves and he named it **X-rays**.

X-rays are produced when high-speed electrons, accelerated through a large potential difference, collide with a metal target and are slowed down. The target is contained within an evacuated glass tube known as an X-ray tube (See Fig. 16).

The production of X-rays can be seen as the **inverse process** of the photoelectric effect. In X-ray production, an incoming electron causes a photon to be emitted. In the photoelectric effect, an incoming photon causes an electron to be emitted.

<i>Photoelectric emission</i>	<i>X-ray production</i>
Transformation of the <i>energy of a photon</i> into the <i>kinetic energy of an electron</i> .	Transformation of the <i>kinetic energy of an electron</i> into the <i>energy of a photon</i> .
Typical energy of photon: a few electron-volts (eV).	Typical energy of electron: in the kilo electron-volts (keV) range.

A typical plot of X-ray intensity versus wavelength is shown in Fig. 17. It consists of sharp peaks or lines superimposed on a broad continuous spectrum. The sharp peaks are called **characteristic lines** or **characteristic X-rays** because they are **characteristic of the target metal used**. The broad continuous spectrum is referred to as **bremsstrahlung** (German for "braking radiation") and is emitted when the bombarding electrons decelerate upon hitting the target.

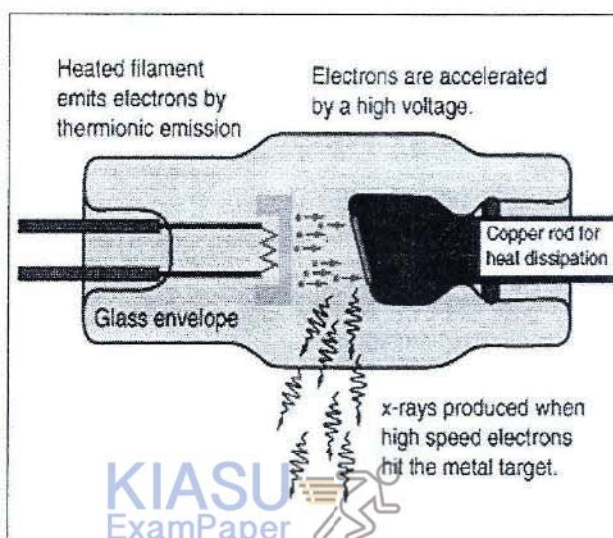


Fig. 16 X-ray tube

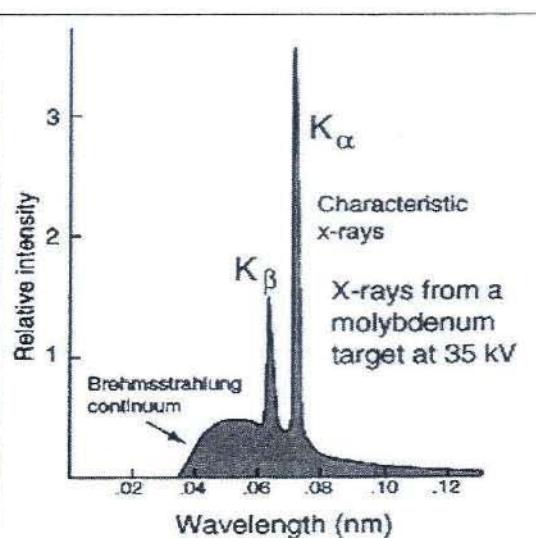


Fig. 17 X-ray spectrum

17.3.2 Explanation of X-ray Emission Spectra

Two distinct processes are involved in X-ray emission (corresponding to two distinct features of the X-ray spectrum in Fig. 17):

- Process 1 accounts for the **bremsstrahlung**,
- Process 2 accounts for the **characteristic lines**.

Explaining the bremsstrahlung

In the first process, some of the electrons pass through the electric field of the atomic nuclei in the metal target. As a result, part or all of the kinetic energies of these electrons are converted directly to a continuous spectrum of photons, including X-rays. The electrons are slowed down or stopped by the target as a result.

The intensity of the continuous background radiation decreases gradually for increasing wavelength, but is cut off sharply at the other end of the spectrum with a minimum wavelength λ_{min} (Fig. 17)

Using the quantum theory of light, one can explain the existence of λ_{min} . When the bombarding electrons interact with the metal atoms in the target, they can lose any fraction of their initial energy, up to the maximum in a single interaction. If the electron loses all its energy in a single interaction, based on quantum theory, a single quantum of electromagnetic radiation (photon) is produced with maximum energy, and hence minimum wavelength.

If the accelerating voltage of the X-ray tube is V , and an electron has charge $-e$, the electron gains kinetic energy eV . Thus the most energetic photon (i.e. highest frequency and shortest wavelength) is produced when the entire electron's kinetic energy goes to produce one photon.

Then,

$$\frac{hc}{\lambda_{min}} = eV \Rightarrow \lambda_{min} = \frac{hc}{eV}$$

(For an electron completely stopped in its first interaction with the metal target)



The minimum wavelength λ_{min} in the bremsstrahlung process is *INDEPENDENT* of the target material. It is dependent only on the accelerating voltage.

Explaining the characteristic lines

The second process accounts for the peaks in the X-ray spectrum at certain frequencies or wavelengths (i.e. the characteristic X-rays in Fig. 17).

Electrons in atoms are grouped in various energy states called 'shells'. The shell nearest to the nucleus is known as the **K** shell (principal quantum number $n = 1$). This is followed by the **L** shell ($n = 2$), the **M** shell ($n = 3$), the **N** shell ($n = 4$) and so on.

In the X-ray tube, highly energetic electrons which have been accelerated by tens to hundreds of kilovolts of potential difference bombard the metal target. These highly energetic bombarding electrons can eject electrons from the *inner shells* (i.e. closer to the nucleus) of the atoms of the metal target. The vacancies created will be quickly filled by electrons dropping down from higher levels, emitting X-rays with sharply defined frequencies associated with the difference between the atomic energy levels of the target atoms.

If E is the difference in energy between the two shells where the electron transition occurs, the frequency f of the radiation emitted is related to E by

$$E = hf.$$

As E is very high for electronic transitions between the K, L and M shells, the frequency f of the emitted radiation is also very high and the wavelength correspondingly short, typically of the order of 10^{-9} m or less (i.e. in the X-ray region of the EM spectrum).

The K series of X-ray lines, K_{α} and K_{β} , is due to the transition of electrons from the L and M shells respectively to the K shell (Fig. 18(a)). There can similarly be L lines produced from transitions to the L shell (Fig. 18(b)).

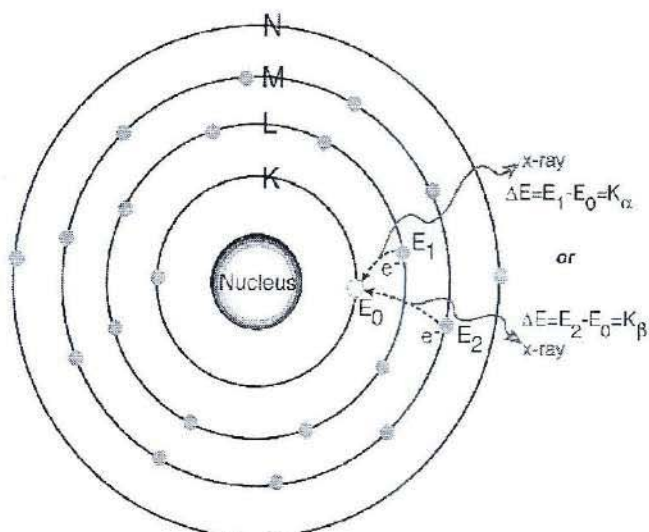


Fig. 18(a)

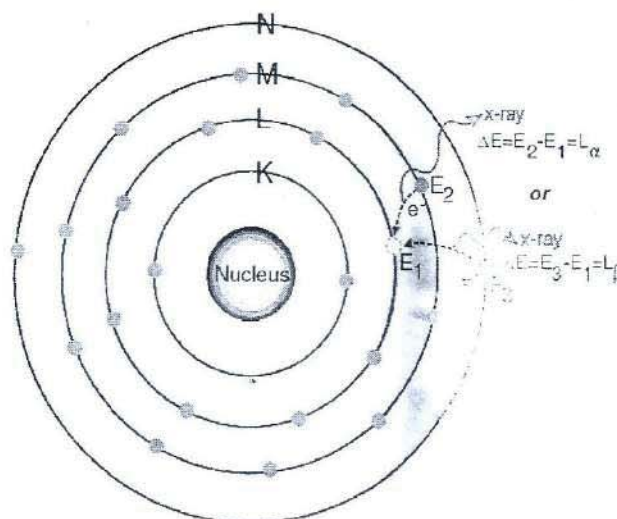
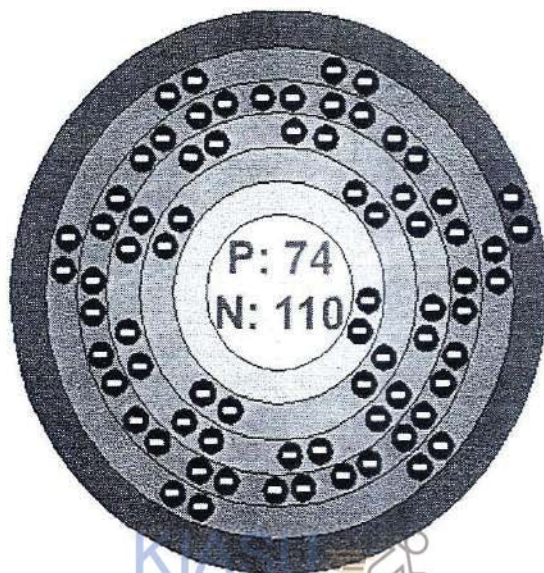


Fig. 18(b)

In Tungsten (74 protons) as seen in Fig. 18(c) and Fig. 18(d), if a K-shell was filled by an L-shell electron, the energy of the emitted radiation is $69.5 \text{ keV} - 10.2 \text{ keV} = 59.3 \text{ keV}$ and it is 69.0 keV if it was filled by an N-shell electron.



<http://www.chemicalelements.com/ban/b0074.gif>

Fig. 18(c)

Occupancy & Shell	Energy (keV)
2 P	0.02
12 O	0.06
32 N	0.50
18 M	2.5
8 L	10.2
2 K	69.5

<http://upload.wikimedia.org/wikipedia/commons/4/42/WBingEnergies.jpg>

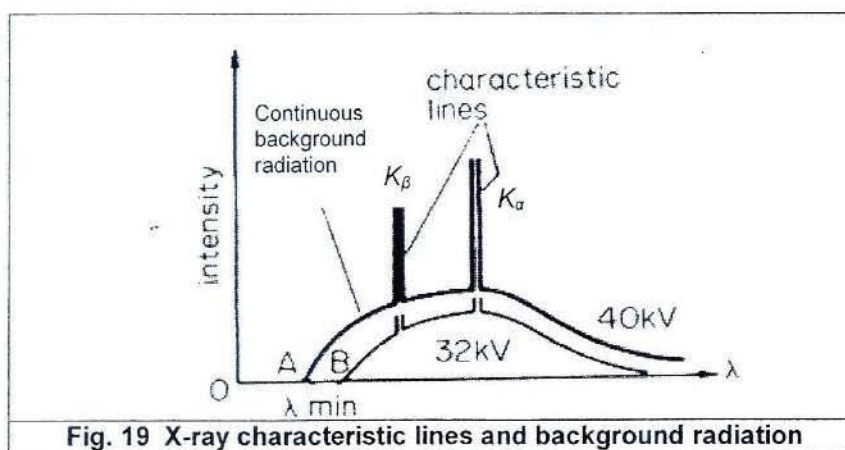
Fig. 18(d)

Thus, **energy changes (de-excitation) in electrons close to the nucleus of metals** result in the characteristic X-rays. In contrast, the emission and absorption spectra we studied in section 17.2 are due to energy changes in the outermost shells of the atom. The energy changes producing optical spectra are about 1000 times smaller. So the frequencies of the optical lines are about 1000 times smaller, that is their wavelengths are about 1000 times longer than those of X-rays.



Unlike the minimum wavelength, the characteristic lines are *DEPENDENT* on the target material.

Fig. 19 shows the characteristic lines, K_α and K_β of a metal and the continuous background radiation (Bremsstrahlung) for two values of accelerating potential, 40 kV and 32 kV, for the X-ray tube (same metal target).



Two observations can be made from Fig. 19:

- 1) The wavelengths of the characteristic lines are independent of the potential difference applied – they are characteristic of the metal.
- 2) λ_{min} for 40 kV spectrum is smaller than that for the 32 kV – this is to be expected as photons produced by electrons accelerated through a potential difference of 40 kV would have a greater maximum possible energy than that produced by electrons accelerated through a potential difference of 32 kV.



This YouTube video answers the question: "How does an X-ray Tube Work?"



17.4 Wave-Particle Duality and the Uncertainty Principle

This video gives the first part of the answer to the question: "What is the Wave/Particle Duality?"



17.4.1 Introduction

We have seen evidence that light can behave like a particle. Yet light also exhibits wave-like properties like diffraction, interference and polarization. Experiments like the Young's double slit experiment clearly demonstrated so.

Can Einstein's photon hypothesis explain the interference and diffraction of light? Unfortunately it cannot. Such behaviour can only be explained using the wave theory of light.

Light behaves **both** as photons and as a wave.

In fact, the equation $E = hf$ contains this contradictory nature of light. The energy of a photon, a particle of light, depends on the frequency of the light. Light can have a frequency only if it is a wave.

17.4.2 de Broglie's Hypothesis – Matter Waves

As a graduate student in 1923, Louis de Broglie made an astounding suggestion. Since light can exhibit particle-like behaviour, then the converse should be true. Particles (like electrons) should also exhibit wave-like behaviour.

de Broglie boldly proposed that a relationship between momentum (of the particle) and wavelength (of the associated matter wave) exists for **all particles**:

$$\lambda = \frac{h}{p}, \quad \text{where } p \text{ is the momentum of the particle}$$

λ is known as the **de Broglie wavelength** of the particle.

Example 9

What is the de Broglie wavelength of an electron that has been accelerated from rest through a potential difference of 54 V?

$$KE = \frac{p^2}{2m} = eV$$

$$\lambda = \frac{h}{p}$$

$$= \frac{h}{\sqrt{2meV}} = \frac{6.63 \times 10^{-34}}{\sqrt{2 \times 9.11 \times 10^{-31} \times 1.6 \times 10^{-19} \times 54}} = 1.6 \times 10^{-10} \text{ m}$$

17.4.3 Evidence of Wave-like Nature of Particles

To exhibit diffraction with light, we had to make very narrow slits that are approximately equal to the wavelength of light (10^{-7} m). For his experiment, Young used a razor blade to scratch slits on a piece of glass that had been blackened with soot. In order to validate de Broglie's hypothesis, we must be able to observe wave-like behaviour with particles (e.g. electrons).

The de Broglie wavelength of electrons is determined to be even shorter. In fact, it is about the same order of magnitude as the size of a typical atom (about 10^{-10} m). This holds the key to showing diffraction of electrons. Diffraction experiments were successfully conducted using a metal crystal. The regularly arranged atoms of the crystal acted as a diffraction grating for the electron "waves". A number of important experiments conducted in the 1930s were based on this principle and clearly demonstrated the wave-like nature of moving electrons.

Davisson and Germer directed a beam of electrons onto a crystal of nickel. The atoms of nickel acted like a diffraction grating and it was observed that the electrons exhibited diffraction behaviour.

Further experiments to demonstrate the wave nature of electrons were conducted by G.P. Thomson. He targeted a mono-energetic beam of electrons at a thin metal foil (Fig. 20). A photographic plate placed behind the foil revealed a bright central maximum surrounded by diffraction rings (Fig. 21).

G.P. Thomson is the son of J.J. Thomson, who discovered the electron. The father was awarded a Nobel Prize for demonstrating that the electron was a particle. The son was awarded one for showing that electrons can behave like waves!

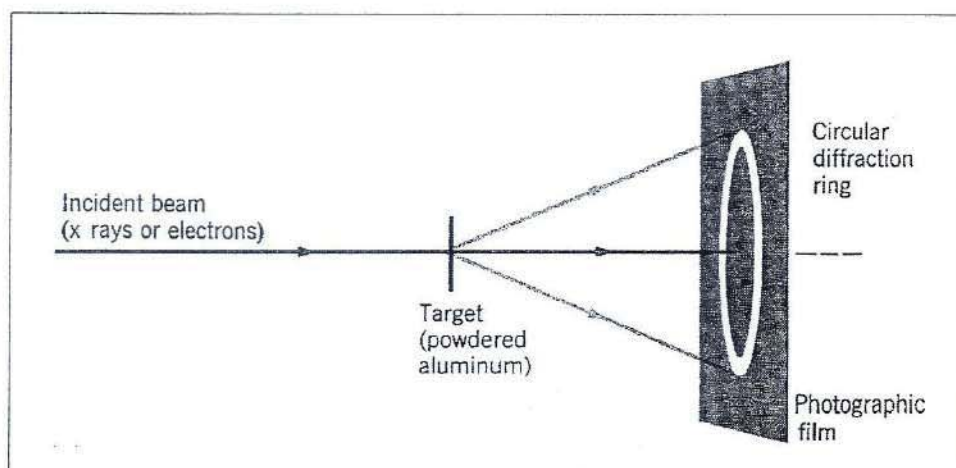


Fig. 20 Set-up used by G. P. Thomson



Fig. 21 Results from Thomson's Experiment

These electron diffraction experiments clearly show that electrons can behave like waves, and detailed calculations show that de Broglie's equation is valid. The wave nature of electrons is something that we now take for granted, and has been put to practical use in devices like the electron microscope.

We conclude that matter or radiation cannot be adequately described by the wave or particle model alone. The wave and particle descriptions are complementary, even though in explaining any particular phenomenon we just need to invoke one model.

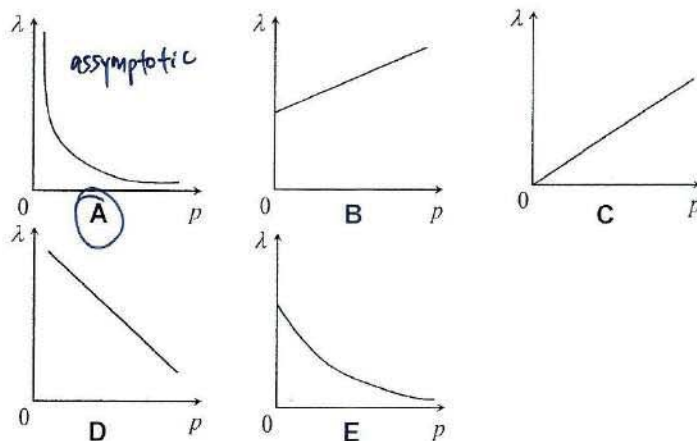
Example 10 (J77/II/33)

Calculate the de Broglie wavelength of a rifle bullet of mass 0.02 kg moving at a speed of 300 m s⁻¹.

$$\begin{aligned}
 p &= (0.02)(300) = 6 \text{ N s} \\
 \lambda &= \frac{h}{p} \\
 &= \frac{6.63 \times 10^{-34}}{6} = 1.11 \times 10^{-34} \text{ m}
 \end{aligned}$$

Example 11 (J80/II/37)

In 1923, de Broglie suggested that an electron of momentum p has properties corresponding to a wave of wavelength λ . Which one of the following graphs correctly shows the relationship between λ and p ?



$$\lambda = \frac{h}{p}$$

17.4.4 The Heisenberg Uncertainty Principle

Another YouTube video, which tries to answer the question: "What is the Uncertainty Principle?"



When we take measurements in an experiment, there will always be experimental uncertainties or errors involved no matter how good the measuring apparatus are. However, according to classical mechanics, there is no fundamental barrier imposed by nature to an ultimate refinement of the apparatus or experimental procedures. That is, the uncertainty in a measurement can be made indefinitely small by using more precise instruments.

Quantum theory predicts, however, that such a barrier does exist and there is a **definite limit** to the precision of any measurement. This limit is **not** due to the limitation of the measuring instrument; rather, it is **inherent in nature** as a result of the **wave-particle nature of matter**.

Imagine generating a wave in a string by periodically shaking it up and down. One would not be able to precisely point out where the wave is along the string, but could certainly give a value for the wavelength, and hence momentum.

On the other hand, if one were to give the string a sudden jerk, one would get a relatively narrow bump travelling down the string. In this case, the location of the wave along the string is more defined, while the wavelength could not be stated (it is not even periodic). For intermediate cases, one could say that it has a fairly defined wavelength and a fairly defined position, but there is an inescapable trade-off between the momentum and position.

In 1927, Werner Heisenberg formally introduced this notion, which is now known as the **Heisenberg Uncertainty Principle**:

$$\Delta x \Delta p_x \geq h$$

If a measurement of position of a particle is made with uncertainty Δx and a simultaneous measurement of linear momentum is made with uncertainty Δp_x , then the product of the two uncertainties can never be smaller than h .

In general terms, the uncertainty principle states that it is impossible to know the exact values of two **related** physical quantities, such as the position and momentum of a particle, **at the same time**.

Example 13

The speed of an electron is $5.00 \times 10^3 \text{ m s}^{-1}$ to an uncertainty of 0.00300%. Find the inherent uncertainty in determining the position of this electron.

$$\Delta p = (9.11 \times 10^{-31}) (5 \times 10^3) \left(\frac{0.003000}{100} \right)$$

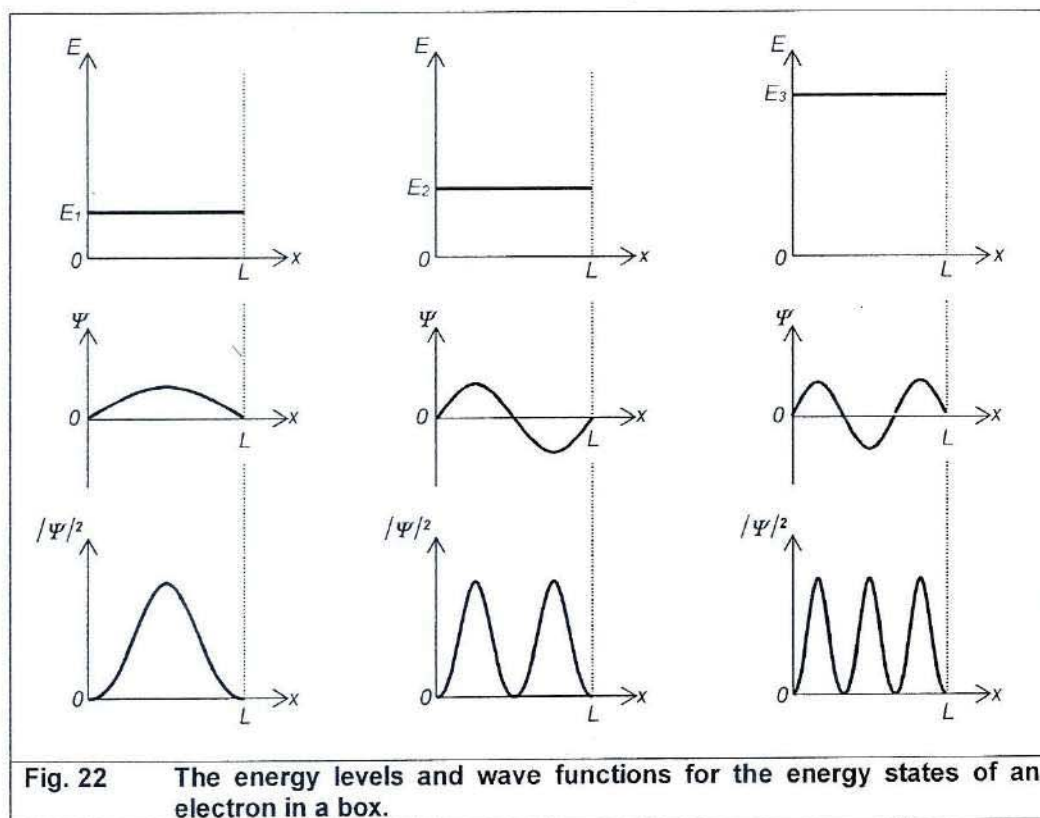
$$= 1.37 \times 10^{-31} \text{ N s}$$

$$\Delta x = \frac{h}{\Delta p} = \frac{6.63 \times 10^{-34}}{1.37 \times 10^{-31}} = 0.0048 \text{ m}$$

Appendix Simple Analysis of an Electron in a Box

Consider the case of an electron trapped in a box (i.e. a potential barrier with well-defined (infinite) boundary). The potential energy of the electron in the box has a constant value, which can be taken to be zero. The electron is confined within the box because it is surrounded by a region of very high potential energy, which the electron cannot enter without breaching the principle of energy conservation.

To further simplify the problem, the electron is confined to one-dimensional motion. That is, the electron only moves along a particular direction in space. As the electron cannot escape from the box the probability of finding it outside the box is zero. If we consider the very edge of the box, the probability of finding the particle at that point can have only one value, so the fact that it is zero outside the box means that it must be zero just inside. The wave functions of this electron can be obtained by solving the Schrödinger equation. Fig. 22 shows three possible wave functions for this electron, each representing a different energy state.



This is analogous to the case of the formation of stationary waves for a string fixed at both ends.

The wavelengths associated with an electron in a box of length L must have one of the values

$$\lambda_n = 2L / n, \text{ where } n \text{ is a whole number}$$

Thus the wavelength of the matter wave is restricted to one of the values corresponding to a whole number of half-wavelength fitting into the box.

By de Broglie's postulate, the magnitude of the electron momentum must have one of the values

$$p_n = h / \lambda_n = nh / 2L$$

Thus the electron momentum is also restricted to a particular set of values. As the potential energy is assigned to be zero and the electron's kinetic energy depends on its mass and its momentum, the total energy is similarly confined to one set of particular values. (i.e. the energy is **quantised** into levels.)

Energy levels of the electron in a box are

$$E_n = \frac{1}{2} m_e v_n^2 = p_n^2 / 2m_e$$

Note: In the case discussed above, we considered matter waves associated with particles trapped in a one-dimensional box. The particle moves in a region where the potential energy is constant. As a result, the momentum and speed of the particle is also constant within the region. But if we considered a particle moving through a region where the potential energy varies, its speed and hence the wavelength associated with the wave function will be changing.

In general, the analysis of a situation where the potential energy varies will require the study of the Schrödinger wave equation. In the case discussed above, where the potential is uniform, the solutions of the Schrödinger equation have the form of standing waves and justified the simplified approach we took.

Content

- The nucleus
- Isotopes
- Nuclear processes
- Mass defect and nuclear binding energy
- Radioactive decay
- Biological effect of radiation

Learning Outcomes

Candidates should be able to:

- infer from the results of the Rutherford α -particle scattering experiment the existence and small size of the nucleus.
- distinguish between nucleon number (mass number) and proton number (atomic number).
- show an understanding that an element can exist in various isotopic forms each with a different number of neutrons.
- use the usual notation for the representation of nuclides and represent simple nuclear reactions by nuclear equations of the form ${}^{14}_7\text{N} + {}^4_2\text{He} \rightarrow {}^{17}_8\text{O} + {}^1_1\text{H}$
- show an understanding of the concept of mass defect.
- state and apply to problem solving the concept that nucleon number, proton number, energy and mass are all conserved in nuclear processes.
- recall and apply the equivalence relationship between energy and mass as represented by $E = mc^2$ to solve problems.
- show an understanding of the concept of binding energy and its relation to mass defect.
- sketch the variation of binding energy per nucleon with nucleon number.
- explain the relevance of binding energy per nucleon to nuclear fusion and to nuclear fission.
- show an understanding of the spontaneous and random nature of nuclear decay.
- infer the random nature of radioactive decay from the fluctuations in count rate.
- show an understanding of the origin and significance of background radiation.
- show an understanding of the nature of α , β and γ radiations. (knowledge of positron emission is not required.)
- show an understanding of how the conservation laws of energy and momentum in β decay were used to predict the existence of the neutrino (knowledge of antineutrino and antiparticles is not required.)
- define the terms activity and decay constant and recall and solve problems using $A = \lambda N$.
- infer and sketch the exponential nature of radioactive decay and solve problems using the relationship $x = x_0 e^{-\lambda t}$ where x could represent activity, number of undecayed particles and received count rate.
- define half-life.
- solve problems using the relation $\lambda = \frac{\ln 2}{t_{1/2}}$
- discuss qualitatively the effects, both direct and indirect, of ionizing radiation on living tissues and cells.

18.1 Introduction

The question of what the smallest component is from which our physical world might be constructed has always been a driving motivation for natural philosophers and scientists in their quests to uncover the constituent fabric of nature.

o The Beliefs of the Early Greeks

- Early Greeks believed that all matter was composed of only four 'elements' - Fire, Earth, Water and Air.
- The Greek philosopher, Leucippus and his student, Democritus were the first to propose that matter is made up of indivisible particles that they called *atomos*, a word derived from *a*-(not) and *tomos* (cut or derived). These ideas were however not readily accepted, with even great intellectual minds like Aristotle and Plato subscribing to the 4 elements belief.

o Modern Atomic Theory - John Dalton (1766 - 1844)

- Early 19th century, John Dalton used *scientific methods* and discovered that many chemical phenomena could only be explained if elements were made up of atoms which he proposed to be basic, indivisible building blocks of matter.

o The Electron

- By 1909, Milikan had completed his first measure of the elementary charge. In 1910, J.J. Thomson discovered the electron.
- These and other experiments (e.g. Rutherford's alpha-particle scattering experiment) showed that all the mass of an atom had to be associated with the *positive charge* (neutrons were not discovered at that time), and not with the electrons.
- It was also known that the overall size of atoms is of the order 10^{-10} m and all atoms except hydrogen contain more than one electron.

o Thomson's Plum Pudding Model



Figure 1.
Thomson's Plum Pudding Model



Figure 2.
A plum pudding

- At that time, what was *not known* was how the mass and charge were distributed within the atom.
- In Thomson's "*Plum Pudding Model*", he proposed:
 - the atom consisted of a sphere of positive charge, of the order of 10^{-10} m in diameter. (like the pudding)
 - electrons were embedded in the positively charged sphere. (like bits of plum in a more or less spherical muffin)

18.2 Rutherford's Alpha-Particle Scattering Experiment

Thomson's plum pudding model of the atom was disproved in 1910 - 1911 by Ernest Rutherford and two of his students, Hans Geiger and Ernest Marsden, at the University of Manchester in England.

18.2.1 Details of Rutherford's Experiment

- **Aim:** To probe the internal structure of the atom.

- **Description**

In Rutherford's alpha(α)-particle scattering experiment, α -particles (Helium nuclei) from a radioactive source were directed towards a thin sheet of gold foil in a vacuum chamber.

- The *vacuum* is necessary because alpha particles only have a short range of about 5 cm in air, and they get scattered before hitting the foil or the screen.
- The gold foil, being very ductile, is rolled into a very thin sheet of 0.00004 cm (about 100 atoms) thick, so that the high energy alpha particles can pass through instead of being absorbed.
- The angular deflection of alpha particles is measured using a microscope that has a glass screen coated with zinc sulphide (Figure 3). A charged particle impacting with the screen will produce a tiny flash of light, called a **scintillation**.
- To determine how the alpha-particles are scattered, the number of scintillations N per minute is obtained for each angular deflection, θ (Figure 4).

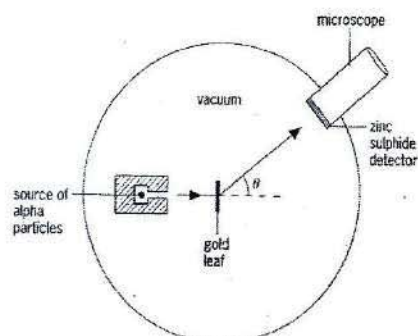


Figure 3.
Set up of Rutherford's Scattering Experiment.

- **Predicted Alpha-scattering based on Thomson's Model (Hypothesis)**

- The positive charge and the negative electrons are distributed throughout the whole atom. Hence, the electric field inside the atom should be quite small.
- Alpha particles are doubly-charged helium ions which are more massive than protons, and fast-moving with a considerable amount of energy, the electrical force on an α -particle that enters the atoms should be quite weak, so the maximum deflection θ observed should only be at most a few degrees.

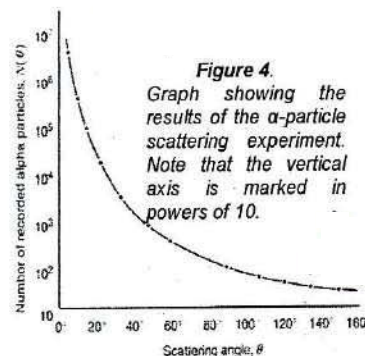


Figure 4.
Graph showing the results of the α -particle scattering experiment. Note that the vertical axis is marked in powers of 10.

- **Actual Observation**

- More than 99% of α -particles passed through with little or no angular deflection,

it is expected to be 100% i.e. for all of the particles if the plum pudding model was correct.

- A few α -particles (~1 in 8000) back-scattered, i.e. suffered deflections of more than 90° . A very small number were deflected backwards.

These results were so unexpected that Rutherford later wrote:

"It was quite the most incredible event that ever happened to me in my life. It was almost as incredible as if you had fired a 15-inch shell at a piece of tissue paper and it came back and hit you."

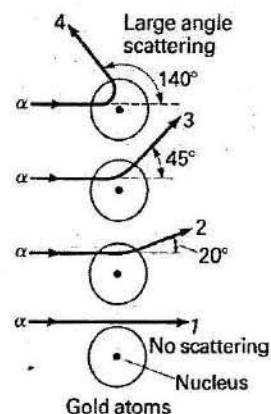


Figure 5. Scattering of α -particles by the nucleus.

18.2.2 Interpretation of Results

Rutherford concluded from the α -particle scattering experiment that:

- i. Very few particles were deflected from their path, i.e. most particles passed through undeflected, indicating that **the positive charge of the atom occupies very little space. Most of the space inside the atom is therefore empty.**
- ii. A very small fraction of α -particles were deflected by very large angles, indicating that **all the positive charge and mass of the gold atom were concentrated in a very small volume within the atom.**

From the data, he also calculated that the radius of the nucleus should be less than 10^{-14} m, at least 10^4 times less than the radius of the atom. (Today, it is deduced that typical size of a nucleus is about 10^{-15} m, i.e. 10^5 times less than the size of the atom.)

He worked out a detailed formula for the scattering, which matched the experimental data to high precision and later proposed a new structure for the atom.

18.2.3 Rutherford's Nuclear Model of an Atom¹

On the basis of his experiment, Rutherford put forward the nuclear model of an atom, which had the following features (Figure 6.):

- (i) There is a positively charged centre in an atom called the nucleus. Nearly all the mass of an atom resides in the nucleus.
- (ii) The electrons revolve around the nucleus in well-defined orbits.
- (iii) The size of the nucleus is very small as compared to the size of the atom.



Rutherford Scattering

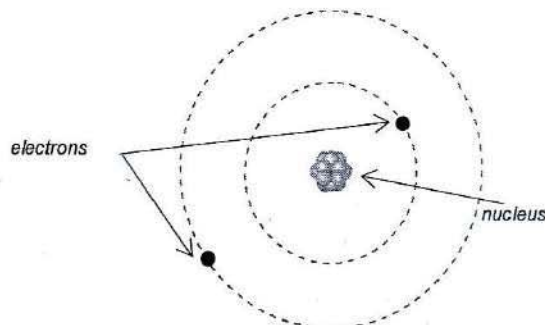


Figure 6. Atomic structure proposed by Rutherford.

¹ Drawbacks of Rutherford's model of the atom:

- (i) It does not obey the Maxwell theory of electrodynamics which dictates that "an accelerated charged particle continuously emits energy". If an electron does so, it should continuously lose its energy and set up a spiral motion, ultimately falling into the nucleus. However, atoms are observed to be quite stable.
- (ii) It could not explain the line spectra of hydrogen atom and its discontinuous spectrum nature.

In order to overcome the objections raised against Rutherford's model of the atom, Neils Bohr proposed the following model of an atom.

- (i) Only certain special orbits known as discrete orbits of electrons, are allowed inside the atom.
- (ii) While revolving in discrete orbits, the electrons do not radiate energy. These orbits or shells are called energy levels. Energy levels in an atom are represented by the letters K, L, M, N, ... or the numbers, $n = 1, 2, 3, 4, \dots$

Example 1: Rutherford Experiment

Young and Freedman 11th Edition - Pg 1461 Example 38.5

An α -particle is aimed directly at a gold nucleus. An α -particle has two protons and a charge of $+2e$, while a gold nucleus has 79 protons and a charge of $+79e$.

What minimum kinetic energy must the alpha particle have to approach within 5.0×10^{-14} m of the centre of the gold nucleus?

Assume that the gold nucleus, which has about 50 times the rest mass of an alpha particle, remains at rest.

By COE,

$$KE_i + PE_i = KE_f + PE_f$$

$$KE_i = \frac{Q_1 Q_2}{4\pi\epsilon_0 d}$$

$$= 7.27 \times 10^{-13}$$

$$= 4.5 \text{ MeV}$$

Example 2: Size of a Nucleus

Akrill, Bennet and Millar (3rd Edition) - Pg 159 Prob. 14.57

- Estimate the size of the full stop at the end of this sentence. 10^{-3} m
- If this dot represents the nucleus of a gold atom, how large, on this scale, would the atom be?
- What is in the space around the nucleus?

b) Diameter of atom $\sim 10^{-10}$ m

Diameter of nucleus $\sim 10^{-14}$ m

Diameter of atom = $10^{-3} \times 10^5 = 100$ m

- c) Mainly empty space w/ e^- orbiting and e^+ nucleus in this space.

KIASU
ExamPaper



Lecture example
solutions

18.3 Some Properties of the Nucleus

18.3.1 Basic Building Blocks of the Nucleus

The basic building blocks of the nucleus are the **proton** and **neutron**:

- Experiments show that we can model a nucleus as a sphere with a radius R that depends on the total number of **nucleons** (neutrons and protons) in the nucleus (Figure 7).
- In a **neutral atom**, the **no. of electrons and protons are the same**.
- The masses and charges of these particles are given below:

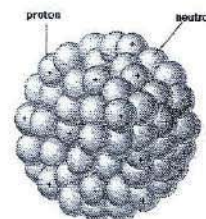


Figure 7. The nucleus is approximately spherical and contains protons & neutrons.

	Proton, ${}_1^1p$	Neutron, ${}_0^1n$	Electron, ${}_{-1}^0e$
	m_p	m_n	m_e
Mass	1.007276 u (1.672622×10^{-27} kg)	1.008665 u (1.674927×10^{-27} kg)	0.000548580 u (9.10938×10^{-31} kg)
Charge	+ e	0 (neutral)	- e

Electronic charge $e = 1.60 \times 10^{-19}$ C

Unified atomic mass, $u = 1.66 \times 10^{-27}$ kg

- From the above table, we can see that over 99.9 % of the mass of an atom comes from its nucleus.

18.3.2 Nuclide Notation

- A **nuclide** is a species of atom characterized by the specific constitution of its nucleus, i.e., its **number of protons** and its **number of neutrons**.
- The **proton number, Z** , is the number of protons in the nucleus. (Z is also called the **charge number**, or **atomic number**);
- The **nucleon number, A** , is the number of **nucleons**. A nucleon is either a neutron or proton in the nucleus. (A is also known as the **mass number**)
- Therefore, the number of neutrons in the nucleus is $N = A - Z$
- The **notation** for a nuclide X with proton number Z and nucleon number A is ${}_Z^AX$.

18.3.3 Isotopes

- Isotopes** are atoms that have the same number of protons but different number of neutrons.
- Examples of isotopes of hydrogen : ${}_1^1\text{H}$ (hydrogen); ${}_1^2\text{H}$ (deuterium); ${}_1^3\text{H}$ (tritium)
- Examples of isotopes of carbon:

	Abundance in Nature	Remarks
${}_6^{11}\text{C}$	Trace	Radioactive : Half-life of 20.4 min
${}_6^{12}\text{C}$	98-99 %	Stable
${}_6^{13}\text{C}$	1-10 %	Stable
${}_6^{14}\text{C}$	Trace	Radioactive : Half-life of 5730 yrs

- Isotopes have nearly identical chemical properties (same proton number) but different physical properties (different nucleon number).

18.3.4 Common Units used in Nuclear Physics

- **Atomic Mass:** Atomic masses are usually expressed in terms of the unified atomic mass unit (u).
- **Atomic Mass Unit (u):** One unified atomic mass unit is one-twelfth the mass of the carbon-12 atom.

$$\text{Hence, } 1 u = \frac{1}{12} \left(\frac{\text{mass of 1 mole of carbon-12}}{\text{Avogadro's number}} \right) = \frac{1}{12} \left(\frac{0.012 \text{ kg}}{6.02 \times 10^{23}} \right) = 1.66 \times 10^{-27} \text{ kg}$$

- **Relative Atomic Mass:** It is common to find in data about a particular atom, the relative atomic mass, m_r . The relative atomic mass (m_r) of an atom is the ratio of the mass of the atom to the unified atomic mass unit.

$$m_r = \frac{\text{mass of atom}}{\text{one-twelfth the mass of a } {}^{12}_6\text{C atom}}$$

It follows from the definition that the *mass of an atom expressed in unified atomic mass units is numerically equal to its relative atomic mass* e.g. the relative atomic mass of Iron (Fe) is 55.845, then the mass of one atom of Fe is 55.845 u .

Notes:

- ✓ Relative atomic mass has no unit.
- ✓ The mass number A should not be confused with the atomic mass. Mass number is an integer specific to a nuclide and has no unit – it is simply a count of the number of nucleons. Atomic mass is generally not an integer because it is a weighted average of the masses of the naturally occurring isotopes of a given element, and has a unit of u .
- **The Electron-Volt :** In nuclear physics, the unit of energy called an electron-volt (eV) is often used :

$$1 \text{ eV} = (1.6 \times 10^{-19} \text{ C})(1 \text{ J/C}) = 1.6 \times 10^{-19} \text{ J}$$

Usually, slightly larger units are more convenient, such as:

$$\begin{array}{ll} 1 \text{ keV} = 1 \times 10^3 \text{ eV} & \text{and} \quad 1 \text{ MeV} = 1 \times 10^6 \text{ eV} \\ (= 1.6 \times 10^{-16} \text{ J}) & (= 1.6 \times 10^{-13} \text{ J}) \end{array}$$

18.4 Nuclear Reactions

The nuclei of atoms can change, or be made to change, in various ways. Changes to the nuclei of atoms are called nuclear reactions.

Examples

- **Radioactive decay :**
It is a process where an unstable nucleus spontaneously and randomly changes into a different nucleus, emitting radiation (α , β , γ) as it does so.
- **Nuclear fission :**
Occurs when a large atomic nucleus splits into two smaller nuclei of approximately the same size.
- **Nuclear fusion :**
Occurs when two smaller atomic nuclei join together (fuse) to form a single larger nucleus. This can only occur when the smaller nuclei have very high energy (e.g. at very high temperatures).

18.4.1 Nuclear Reactions and Conservation Laws

Nuclear reactions may be shown in a form similar to chemical equations, for example



The transformations of particles must follow certain conservation laws. In all nuclear reactions, **the following quantities** must be equal on each side of the equation:

- Sum of nucleon numbers
- Sum of proton numbers
- Net electric charge
- Total energy (sum of mass-energy, kinetic energy and em radiation)

Mass-energy



Real meaning of $E=mc^2$

- It is observed that in a nuclear reaction, the total mass of reactant and products are different. A decrease in mass is always accompanied by release of energy, whereas an increase in mass is always accompanied by absorption of energy.
- The energy E associated with a change of mass m is given by the mass-energy equivalence relation: $E = mc^2$ where c is the speed of light.

- Total momentum

Example 3: Nuclear Reaction Equations

Milner SAQ 1.5 and SAQ 1.6

a) The nuclear equation for a particular fission is given by : ${}_{92}^{235}\text{U} + {}_0^1\text{n} \rightarrow {}_{44}^{141}\text{Cs} + {}_{47}^{92}\text{Rb} + 2{}_0^1\text{n}$

Determine i) the nucleon number, and ii) the proton number of the rubidium atom

b) Two deuterium nuclei can fuse to produce a tritium nucleus and a proton. Write a nuclear equation for this reaction.

Answer:

(a) (i) Nucleon number = $235 + 1 - (141 + 2(1)) = 93$

(ii) Proton number = $92 - 44 = 48$

(b) ${}_1^2\text{H} + {}_1^2\text{H} \rightarrow {}_1^3\text{H} + {}_1^1\text{p}$

Example 4: Energy-Mass Equivalence

What is the energy equivalent, in MeV, of one unified atomic mass, 1 u?

Answer:

$$\begin{aligned} \text{Energy of } 1\text{ u} &= 1.66 \times 10^{-27} \times (3.00 \times 10^8)^2 \text{ J} \\ &= \frac{1.494 \times 10^{-10}}{1.60 \times 10^{-19} \times 10^6} = 934 \text{ MeV} \end{aligned}$$



Lecture example solutions

Example 5: Conservation of Mass Energy → Mass Loss

Serway College Physics p958 – main text (modified)

Consider the following nuclear reaction: ${}^2_1\text{H} + {}^{14}_7\text{N} \rightarrow {}^{12}_6\text{C} + {}^4_2\text{He}$

The masses of the nuclei in the above equation are given by:

Mass of ${}^2_1\text{H} = 2.014\,102\,u$ Mass of ${}^{12}_6\text{C} = 12.000\,000\,u$ Mass of ${}^{14}_7\text{N} = 14.003\,074\,u$ Mass of ${}^4_2\text{He} = 4.002\,602\,u$

- What is the mass "lost" by the system in the nuclear reaction?
- What is therefore the equivalent energy "lost" by the system, in MeV?
- What is this lost energy manifested as?
- Suggest why the combined energies of the products are usually higher than the value calculated in (b)?

$$a) (2.014102u + 14.003074u) - (12.000000u + 4.002602u) = 0.014574u$$

$$b) (0.014574)(1.66 \times 10^{-27}) \cdot (3 \times 10^8)^2 = 2.177 \times 10^{-12} \text{ J} = 13.6 \text{ MeV}$$

c) KE of pdt nuclei (& possibly gamma radiation)

d) Nuclear rxns require the rxt nuclei to be brought close tgt. The rxt nuclei must possess non-zero initial KE to do work against the repulsive forces b/w the rxt nuclei

KE of pdt nuclei + final rest energy = initial KE of rxt nuclei + initial rest energy.



Relating BE & MD - ppt slideshow

Example 6: Conservation of Mass-Energy → Mass Gain

(N82/1/11 – modified)

A nitrogen nucleus ${}^{14}_7\text{N}$ bombarded by an alpha particle with a certain kinetic energy, transmutes to an oxygen nucleus ${}^{17}_8\text{O}$ and a proton.

- Write an equation for this nuclear reaction, showing the mass number and atomic numbers of the particles involved.
- Find the minimum energy of the alpha particle to make this reaction occur.
- By also considering the linear momentum of the nuclei, suggest why even if the alpha particle were to possess this minimum energy, it will be unlikely for the reaction to occur.

Given:

Mass of ${}^4_2\text{He} = 4.002\,602\,u$ Mass of ${}^{17}_8\text{O} = 16.999\,133u$ Mass of ${}^{14}_7\text{N} = 14.003\,074\,u$ Mass of ${}^1_1\text{H} = 1.007\,825\,u$ 

$$b) m = [(16.999133 + 1.007825) - (14.003074 + 4.002602)] = 0.001282u$$

$$E = mc^2 = (0.001282)(1.66 \times 10^{-27}) \cdot (3 \times 10^8)^2 = 1.915 \times 10^{-13} \text{ J} = 1.20 \text{ MeV}$$

c) By PCOM, since total init. p was not 0,

the pdt nuclides must also have non-zero p & thus KE.

Initial rest energy + KE of rxt = Final rest energy + KE of pdt nuclei

As total initial momentum > 0 , thus final total momentum > 0 .Final KE of the pdts > 0 (i.e. pdts will have speed)

This means that the pdts possess more mass-energy than can be accounted for by the 1.20 MeV energy possessed by the alpha particle. The violation of the conservation of mass-energy implies that the rxn is unlikely to occur.

18.5 Nuclear Force & Nuclear Binding Energy

An atomic nucleus consists of positively charged protons and neutral neutrons concentrated in a tiny volume in space. Electrostatic repulsion between the positively charged protons should have caused the disintegration of the nuclei spontaneously. However we know that other than radioactive nuclei, nuclei are relatively stable.

It dawned on physicists that there must be another force, other than the electrostatic force, which holds the nucleons together – the Strong Nuclear Force.

18.5.1 Strong Nuclear Force & the Stability of Nucleus

o Nature and Range of the Strong Nuclear Force

- At typical nucleon separation 1.3 fm (1.3×10^{-15} m), it is a very strong attractive force (10^4 N).
- At much smaller separations between nucleons, the force becomes repulsive.
- Beyond about 1.3 fm separation, the force quickly dies off to zero.
- Thus, the nuclear force is a **short-range force**.
- The much smaller Coulomb force between protons has a much longer range and becomes the only significant force between protons when their separation exceeds about 2.5×10^{-15} m.

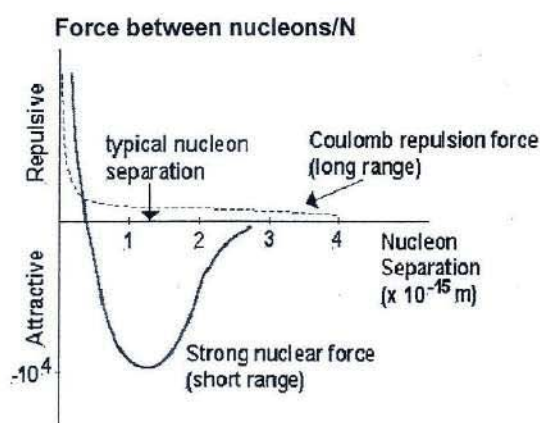


Figure 8. Graph of Strong force between 2 nucleons

o The nuclear force is independent of charge

The nuclear forces associated with the proton-proton, proton-neutron and neutron-neutron interactions are the same.

18.5.2 Stability of a Nucleus

Apart from the nucleus of ${}^1_1\text{H}$ atom, there is a very delicate balance between the strong nuclear forces that act between adjacent nucleons and the electrostatic forces that act between the protons and this determines how **stable** a particular nucleus is.

Out of the 339 naturally occurring nuclides, 254 are stable and 85 unstable. Beyond the 339 naturally occurring nuclides, more than 3000 radionuclides have been artificially produced.

A plot of N vs. Z for a number of stable nuclides is given in Figure 9.

- o Note that for **light nuclides**, they are *most stable* if they contain *almost equal numbers of neutrons and protons* (i.e. $N = Z$).
- o **Heavier nuclides** are *more stable* if $N > Z$.

This difference can be partially understood by recognizing that as the number of protons increases, the electrostatic repulsive force increases, and this tends to break the nucleus apart. Hence, more neutrons are needed to keep the nucleus stable as neutrons are affected only by the attractive nuclear forces.

- o **Beyond $Z = 83$** , there are no stable nuclides. This is because the repulsive forces between the protons can no longer be compensated by the addition of neutrons.

Nuclides that contain more than 83 protons are unstable and tend to decay or disintegrate into other nuclides.

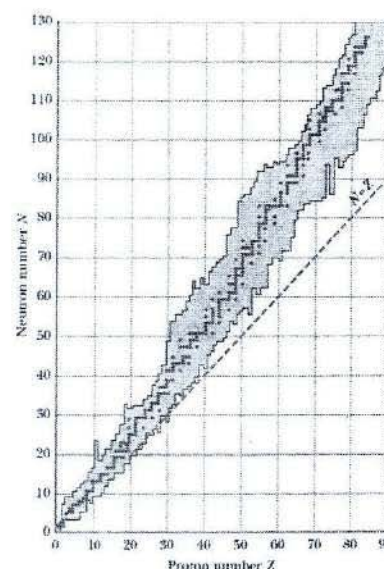


Figure 9.

A plot of the neutron number N versus the proton number Z for the stable nuclides (solid points). The dashed line corresponds to $N = Z$. The plots are centred on the so-called line of stability. The shaded area shows radioactive (unstable) nuclides.

18.5.3 Nuclear Binding Energy

- Similar to the ionisation energy (also known as electron binding energy) that is required to supply to an atom to completely liberate an orbital electron in its ground state (lowest energy level), work has to be done against the strong nuclear force to disassemble a nucleus into individual unbound protons and neutrons. The energy required at nuclear level is in the order millions of times greater than the ionization energy at atomic level.

The **binding energy** of a nucleus is the energy required to separate the nucleus into its constituent neutrons and protons (Figure 10).

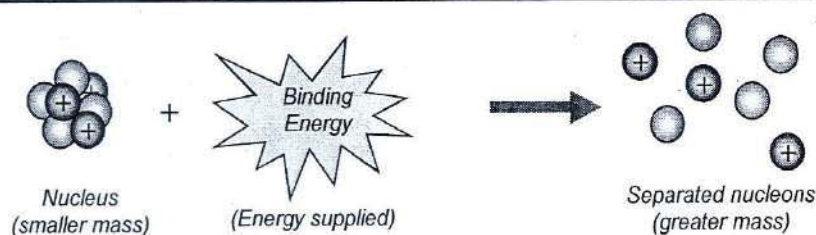


Figure 10. Energy must be supplied to break up a nucleus into its constituent neutrons and protons.

- This binding energy is also the energy that is released when free protons and neutrons are bound together in a nucleus (Figure 11).

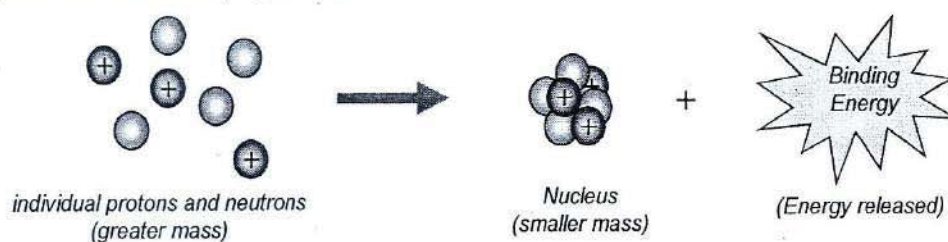


Figure 11. If the nucleons bind together to form a nucleus, energy will be released.

18.5.4 Mass Defect

- Recall that *mass* and *energy* are actually equivalent and are related by Einstein's equation $E = mc^2$
- Since energy is released when the individual protons and neutrons bind to form a stable or semi-stable nucleus, **the nucleus will have a smaller mass than the total mass of the separate neutrons and protons.**

This difference in mass Δm is known as the **mass defect**.

- In general, to calculate the mass defect Δm for a **nucleus** that has Z protons and N neutrons:

$$\Delta m = (\text{Sum of mass of constituent nucleons}) - (\text{Mass of nucleus})$$

$$\Delta m = Zm_p + Nm_n - M_{\text{nucleus}}$$

where m_p : mass of a proton, m_n : mass of a neutron, M_{nucleus} : mass of the nucleus

- It therefore follows from Einstein's mass-energy equivalence relationship that:

$$\text{Binding Energy } E = (\Delta m)c^2$$

where Δm is the mass defect.



BE, Nuclear Strong Force

Example 7: Binding Energy of Deuterium

Serway College Physics p943

The nucleus of the deuterium atom, called the deuteron, consists of a proton and a neutron. Calculate the deuteron's binding energy in MeV, given that its nuclear mass is 2.014 102 u.

Mass of proton $m_p = 1.007276 \text{ u}$

Mass of neutron $m_n = 1.008665 \text{ u}$

$$\Delta m = (1.007276 + 1.008665) - 2.014102 = 1.839 \times 10^{-3} \text{ u}$$

Binding energy of deuteron, $\leftarrow 1 \text{ u}^8$

$$B.E = (1.839 \times 10^{-3} \text{ u}) (934 \text{ MeV}) = 1.72 \text{ MeV}$$

18.5.5 Binding Energy per Nucleon (B.E/A)

- Although the larger an atomic nucleus, the greater its total binding energy, it is the **binding energy per nucleon that determines the stability of the nucleus.**

The binding energy per nucleon represents the average energy required to remove a nucleon from the nucleus.

- Therefore the **greater the binding energy per nucleon, the more stable is the nucleus.**
- A graph of binding energy per nucleon against the number of nucleons (Figure 12) shows several interesting features:
 - Very small nuclei and very large nuclei are both less stable than nuclei of intermediate sizes.
 - The average binding energy per nucleon is about 8 MeV.
 - ^{56}Fe is one of the most stable and abundant nuclei. It resides close to the top of the curve, with a binding energy per nucleon of 8.8 MeV.
 - Very small nuclei tend to form more stable heavier nuclei through **fusion**.
 - Similarly, very large nuclei can be split into smaller, more stable nuclei through **fission**.

Element	Mass of protons & neutrons/ u	Nuclear Mass/ u	Binding Energy / MeV	BE/A / MeV
Deuterium	2.01594	2.01355	2.23	1.12
Helium 4	4.03188	4.00151	28.29	7.07
Lithium 7	7.05649	7.01336	40.15	5.74
Beryllium 9	9.07243	9.00999	58.13	6.46
Iron 56	56.44913	55.92069	492.24	8.7906
Silver 107	107.86187	106.87934	915.23	8.55
Iodine 127	128.02684	126.87544	1072.53	8.45
Lead 206	207.67109	205.92952	1622.27	7.88
Polonium 210	211.70297	209.93683	1645.16	7.83
Uranium 235	236.90849	234.99351	1783.80	7.59
Uranium 238	239.93448	238.00037	1801.63	7.57

Table 12. binding energies per nucleon for some common elements

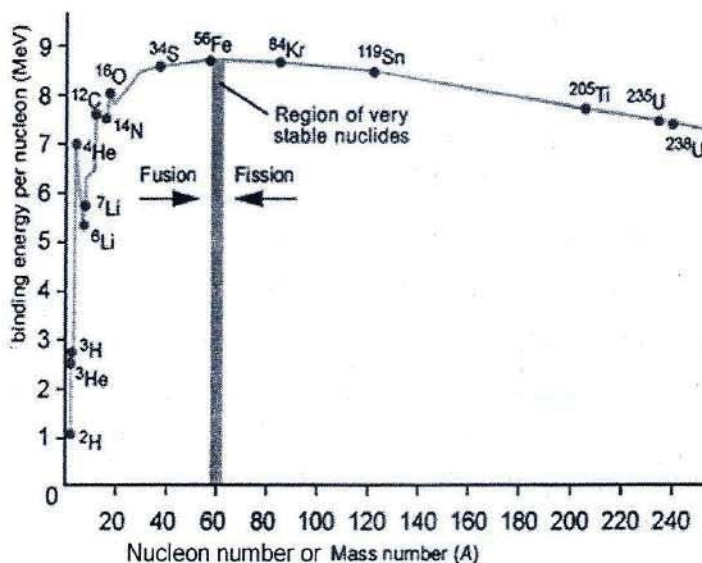


Figure 12. Graph of Binding Energy per nucleon vs the Nucleon Number (A)

Example 8: Binding Energy per Nucleon

TYS N85/III/11 (part)

The conventional notation for a certain nuclide is ${}^{81}_{35}\text{Br}$.

- Write down (i) the number of protons, (ii) the number of neutrons, in its nucleus.
- The mass of its nucleus is 80.8971 u. Taking the proton and neutron masses to be 1.0073 u and 1.0087 u respectively, deduce the binding energy per nucleon.
- Sketch a labeled graph showing how the binding energy per nucleon varies with mass number, indicating the approximate position of ${}^{81}_{35}\text{Br}$ on the curve.
- Explain the implications for the stability of a given nuclide of a low value of binding energy per nucleon.

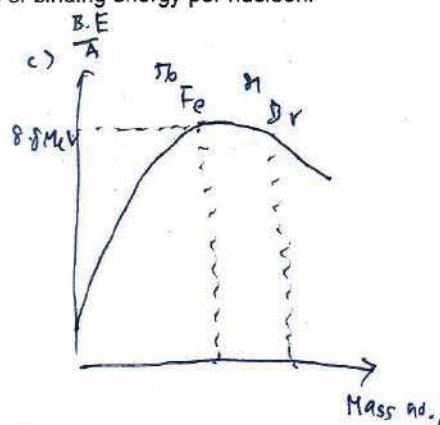
a) (i). No. of protons, $Z = 35$ ii). No. of neutrons, $A - Z = 81 - 35 = 46$

b) Mass defect of Br-81 nucleus,

$$\Delta m = (35 \times 1.0073) + (46 \times 1.0087) - 80.8971 = 0.7586 \text{ u}$$

$$\text{B.E. of Br-81 nucleus} = 0.7586 \times 934 = 708 \text{ MeV}$$

$$\text{B.E. per nucleon} = \frac{708 \text{ MeV}}{81} = 8.74 \text{ MeV}$$



d) B.E. per nucleon represents the avg. energy required to separate a nucleon from the nucleus. Lower value of B.E. per nucleon means less energy required to separate nucleon from the nucleus. Thus a given nuclide of

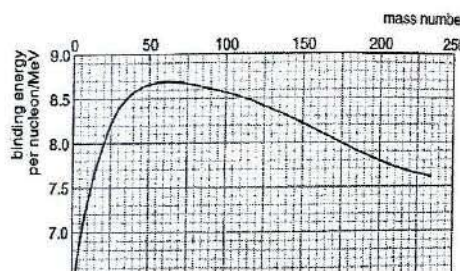
Example 9: Binding Energy per Nucleon

TYS N96/III/6 (part, modified – graph from Akrill, Benett and Millar)

low value of B.E. per nucleon will be less stable

The binding energy per nucleon varies with nucleon number in the way shown in the figure. During one particular fission process, a Uranium-236 nucleus gives, among its fission products, a Strontium-90 nucleus and a Xenon-143 nucleus.

- Use the values on the figure to calculate the energy released during this fission process.
- What other fission particles are produced by this process? Why do these particles not have to be taken into account in the calculation in (a)?
- Why does a release of energy occur when there is an increase in the binding energy?



a) Energy required to initiate the fission rxn

$$(\text{re. B.E. of U-236}) = (7.6)(236) = 1793.6 \text{ MeV}$$

$$\text{B.E. after rxn} = (8.6)(90) + (8.3)(143) = 1960.9 \text{ MeV}$$

$$\text{Net energy released during fission process} = 1960.9 - 1793.6 = 167.3 \text{ MeV}$$

b) ${}^{236}_{92}\text{U} + {}^1_0\text{n} \rightarrow {}^{90}_{36}\text{Sr} + {}^{143}_{56}\text{Xe} + 4 {}^1_0\text{n}$

Neutrons are produced. These need not be taken into account in the calculation in (a) as there are no binding energy involved for single isolated neutron.

c) Total B.E. ↑ (i.e. sum of B.E. of fission nuclide < B.E. of pds), if there is a ↓ in total rest mass (i.e. rest mass of fission nuclide > rest mass of pds)

By the mass-energy equivalence principle, a ↓ in rest mass implies a release of energy (by $E = mc^2$).

18.6 Radioactivity

As we have seen, the stability of a nucleus is determined by its binding energy per nucleon. One of the possible routes which unstable nuclei can become more stable is via radioactive decay.



What is radioactivity

- Radioactive decay is a **spontaneous** and **random** process where an unstable nucleus changes into a more stable nucleus, emitting radiation as it does so.

➤ Spontaneous Process:

When we say that radioactive decay is a **spontaneous** process, it means that radioactive decay is not triggered by external factors or influences, and the rate of decay is not affected by environmental conditions.

➤ Random Process :

When we say that radioactive decay is a **random** process, it means that it is not possible to predict which nucleus will decay or when a particular nucleus will decay, and the probability of decay per unit time is constant and the same for all nuclei in the sample.

- Hence, the decay of a radioactive atom *is not affected by any chemical reaction*. It *can neither be sped up nor slowed down* and is *independent of physical conditions* such as temperature, pressure and the decay of other atoms.

○ Evidence of Randomness of Radioactivity

- The random nature of radioactive decay can be demonstrated by observing the fluctuations in count rate.
- When a Geiger-Muller tube (See *Appendix A*) is placed near a source, the irregularity of the counts and the fluctuations in the count rate recorded by the tube show the randomness of radioactive decay (Figure 13).

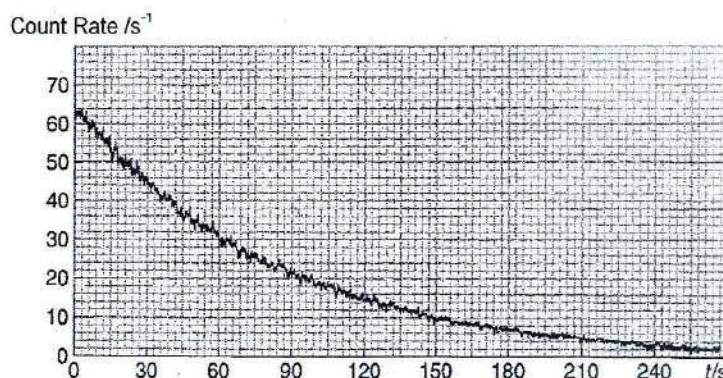


Figure 13. The fluctuations in the count rate of a radioactive decay demonstrate its random nature.

- When one nucleus changes into another through a spontaneous decay the process is known as *transmutation*.

18.6.1 Types of Radioactive Decay

Alpha Decay

- An α -particle is a fast moving Helium-4 (${}^4_2\text{He}$) nucleus emitted from an unstable nucleus. It has a charge of $+2e$.
- Alpha decay occurs when an unstable heavy nucleus (the "parent") ejects an alpha particle, transforming itself into a different nuclide (the "daughter") after losing two protons and two neutrons (Figure 14).



Types of radiation

- Alpha decay can be written symbolically as:

$${}^A_Z\text{X} \rightarrow {}^{A-4}_{Z-2}\text{Y} + {}^4_2\text{He}$$

- where ${}^A_Z\text{X}$ is known as the **parent nucleus** and ${}^{A-4}_{Z-2}\text{Y}$ is known as the **daughter nucleus**.

e.g. ${}^{238}_{92}\text{U}$ is an alpha emitter and decays according to the following scheme (Figure 14):

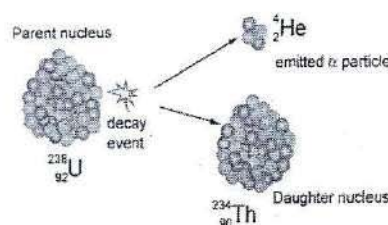
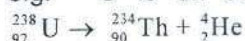


Figure 14. Alpha decay of U-238.

- Alpha emitting nuclei tend to be very heavy nuclei (with very high proton number).

Energy and Momentum in α Decay

Energy is released during an alpha decay. By the mass-energy equivalence principle, the combined rest mass of the daughter nucleus and the alpha particle is smaller than the rest mass of the parent nucleus. The energy released shows up in the **kinetic energy of the daughter nucleus and α -particle**.

Since the total linear momentum is conserved, when the parent nucleus was initially at rest, **the daughter nucleus and the α -particle must have equal but opposite momentum**. Since the α -particle has a much smaller mass, it therefore carries away most of the kinetic energy.

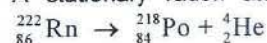
$$KE = \frac{p^2}{2m}$$

(Useful alternative expression of the kinetic energy of a particle:

Example 10: Alpha Decay

TYS N83/1/13 (part)

A stationary radon atom may decay spontaneously into a polonium atom and a helium atom, as shown:



The rest masses of these atoms are: ${}^{222}\text{Rn}$, 222.0176 u; ${}^{218}\text{Po}$, 218.0090 u; ${}^4\text{He}$, 4.0026 u.

- Explain carefully how energy is conserved in this decay. Assuming that no γ -ray is emitted, calculate the total kinetic energy of the decay products.
- Explain carefully how momentum is conserved and calculate the speed of each of the decay products.

a) Total K-E of decay prod's

$$\begin{aligned} E &= (\Delta m)c^2 \\ &= (222.0176 - 218.0090 - 4.0026 \text{ u}) (1.66 \times 10^{-27} \text{ kg}) (3.0 \times 10^8 \text{ m/s})^2 \\ &= 8.964 \times 10^{-13} \text{ J} = 5.603 \text{ MeV} \end{aligned}$$

b) By PCON, $m_a v_a + m_{Po} (-v_{Po}) = 0$ — (1)

$$\text{By conservation of mass-energy, } E = \frac{1}{2} m_a (v_a)^2 + \frac{1}{2} m_{Po} (v_{Po})^2 \text{ — (2)}$$

Solve (1) & (2),

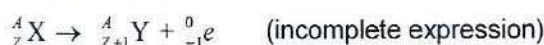
Speed of α particle, $v_a = 1.6 \times 10^6 \text{ m/s}$

Speed of Po -218 atom, $v_{Po} = 2.94 \times 10^5 \text{ m/s}$

Beta Decay

- **Beta decay** is a type of radioactive decay in which a β -particle is emitted.
- A β -particle can be
 - a high speed electron ejected from an unstable nucleus, referred to as β^- particle, or
 - a high speed positron² ejected from an unstable nucleus, referred to as β^+ particle.
(For the purpose of the A-level, β -particle refers to the electron.)
- Beta decay can be :
 - **Beta-minus (β^-) decay**,
where an **unstable neutron in the nucleus transforms into a proton and electron and ejects the electron (β^- particle).** (Figure 15a)

The decay process can be written as :



- **Beta-plus (β^+) decay**,
where an **unstable proton in the nucleus transforms into a neutron and ejects the positron (β^+ particle).** (Figure 15b)

The decay process can be written as :

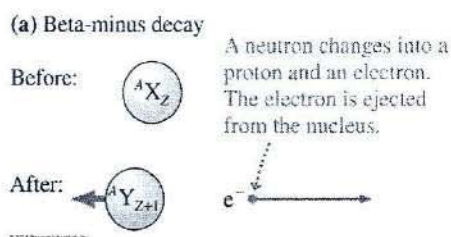
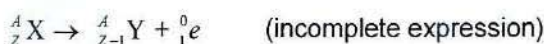


Figure 15(a)

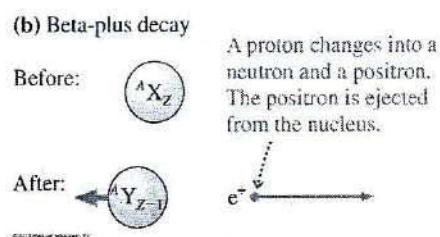


Figure 15(b)

- In both cases:
 1. As the proton number has changed, the parent nucleus **transmutes** into another nucleus
 2. Although the respective numbers of protons and neutrons in the nucleus change during beta decay, the nucleon number remains the same.
 3. Although β^- decay results in an electron being ejected from the nucleus, the electron DOES NOT exist within the nucleus prior to the β -decay. It is created during the radioactive decay process. (This process is NOT ionization, where an orbital electron is freed from an atom.)

² Positron is the antiparticle of the electron. Corresponding to most kinds of particles, there is an associated antiparticle with the same mass and opposite charge.

More on Beta Decay – the Existence of Neutrinos

Consider the β decay or β^+ decay represented above. (Figure 15). If you apply *conservation of linear momentum* and *conservation of mass-energy* to each decay, you will see that for an radioactive isotope A_ZX , the β radiation emitted should have only a single momentum and energy.

However, experimentally, it has been found that beta particles from a single type of nucleus are emitted over a range of energies and momentum, violating both conservation of energy and momentum.

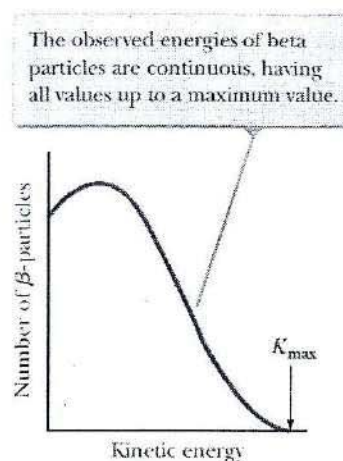
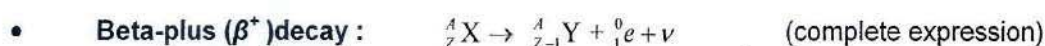
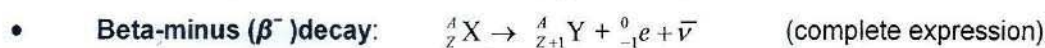


Figure 16

In 1930, Pauli proposed a third particle must be present in the decay products. Enrico Fermi named this particle the *neutrino* (symbol " ν "), meaning the little neutral one.

- **Beta Processes in their most correct and complete form** is therefore:



where $\bar{\nu}$ represents the antineutrino and
 ν represents the neutrino

- The energy emitted in each reaction is shared between the β -particle and the neutrino (or antineutrino in the case of β minus decay). This allows the β particles to be emitted from the nucleus with a range of energies and momentum - ensuring coherence with the conservation of mass-energy and conservation of linear momentum.

- **Properties of Neutrino:**

- From conservation of charge, we can see from the equation above that the neutrino is *neutrally charged*.
- The neutrino also have *negligible mass* as it interacts weakly with other matter and hence have eluded detection for many years.

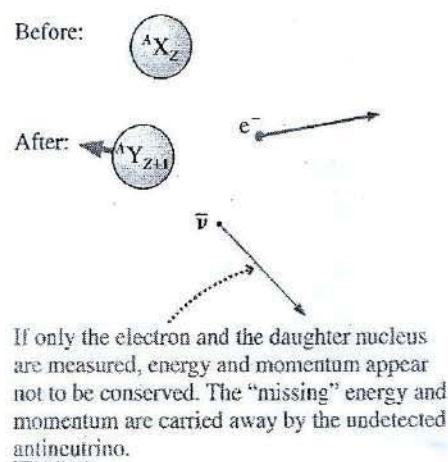


Figure 17

Gamma Decay

- Very often, after an α or β decay, the nucleus is left in an *excited* state.

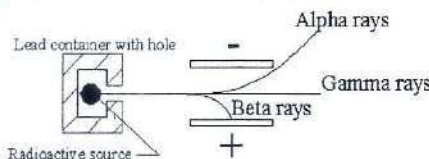
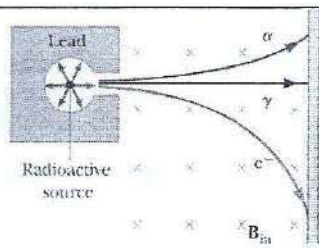
A typical nucleus has a set of allowed energy levels, including a ground state and several excited states. A nucleus in an excited state is unstable and it will undergo decay to a lower energy state by emitting one or more high energy Gamma photons:



(The excited nucleus is denoted by an asterisk *.)

Properties of Radioactive Emissions

The table below shows the properties of the three types of radiations.

	α -particle	β -particle	γ -ray
Nature	${}^4_2\text{He}$ Helium nucleus	${}^0_{-1}e$ Electron (β^- particle)	Photon
Charge	$+2e$	$-e$	0
Ionization	Very strong (producing 10^3 to 10^4 ions per mm of path)	Strong	Weak
Penetrating power	Easily stopped by paper, skin, or a few cm of air	Stopped by ~ 1 m air or 3 mm of aluminum	Reduced significantly by several cm of lead
Deflection by an Electric field	 Figure 18.		
Deflection by a Magnetic Field	 Figure 19.		



Penetrating power of different radiation

18.6.2 Laws of Radioactive Decay

Activity (A) of a Radioactive Decay

The *rate of decay*, or **activity**, of a radioactive sample is the number of disintegrations per unit time.

$$A = -\frac{dN}{dt}$$

where N is the number of radioactive nuclei present.

S.I. unit of activity: s^{-1} or Bq (becquerel)

Notes:

- $\frac{dN}{dt}$ is a negative number, since N decreases with time. The activity is reported as a positive number, since it represents the decrease per unit time.
- $1 \text{ Bq} = 1 \text{ disintegration per second}$.
- Another commonly used unit of activity is **curie (Ci)**, defined as $1 \text{ Ci} = 3.7 \times 10^{10} \text{ Bq}$
The curie is a very large unit. Most of the activity of radioactive substances are measured in millicurie (mCi) or microcurie (μCi).
- The GM tube measures the **count rate (C)** instead of the activity. This is because the detector only captures A PORTION of the radioactive emissions from a sample of radioactive nuclide.

The count rate is actually a fraction of the activity, i.e. $C \propto A$

Background Radiation

- In the absence of specific radioactive source, a radiation detector still picks up a non-zero count rate (typically 20 to 50 counts per minute) from the environment. This measurement is known as the **background count**.
- Background count is due to **ionizing radiation** that is constantly present in the environment. This radiation is known as **background radiation**, and its sources can be natural or man-made.
- Naturally-occurring sources account for about 80% of our exposure, and is contributed by things such as rocks, radon (a colourless and odourless gas that is easily trapped in buildings) and cosmic radiation. Man-made sources include those used in nuclear medicine, nuclear power plants and in consumer products (such as tobacco, which contains radioactive Pb-210). For more information on background radiation, refer to Appendix B.
- **In experiments to study radioactivity, the background count is usually subtracted from the measurement, to eliminate the systematic error in the count rate.**

Decay Constant

Although radioactive decay is a random process, if the number of radioactive nuclei is large, we can associate a probability of decay to the nuclei.

- The rate at which a particular decay process occurs in a sample is proportional to the number of radioactive nuclei present :

$$\frac{dN}{dt} = -\lambda N \dots\dots\dots (1)$$

where λ is the proportionality constant known as the **decay constant**.

Note:

1. Since Activity $A = -\frac{dN}{dt}$, we can also write equation (1) as

$$A = \lambda N \dots\dots\dots (2)$$

2. Equation (1) can also be written as : $\lambda = \frac{(-\frac{dN}{dt})}{N}$

The **decay constant** is the probability per unit time that a nucleus will decay.

3. Rewriting equation (1), we can also have:

$$\frac{dN}{N} = -\lambda dt$$

Integrating, $\int_{N_0}^N \frac{1}{N} dN = -\lambda \int_0^t dt$

$$\Rightarrow \ln\left(\frac{N}{N_0}\right) = -\lambda t$$

$$\Rightarrow \boxed{N = N_0 e^{-\lambda t}} \dots\dots\dots (3)$$

where N is the number of radioactive nuclei present at time t and N_0 represents the number of undecayed radioactive nuclei at $t = 0$.

Equation (3) shows that *the number of undecayed radioactive nuclei in a sample decreases exponentially with time* (Figure 20).

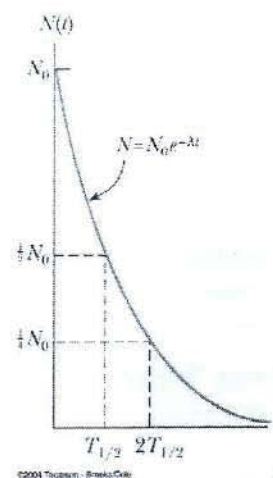


Figure 20. Plot of the exponential decay law for radioactive nuclei. The curve is actually jagged (see Figure 13). The smooth curve is fitted to represent the trend of the decay. The time $T_{1/2}$ is the half life of the sample.

4. Variation of Activity with time

Recall that the activity A of a sample of radioactive nuclei is proportional to the no. of radioactive nuclei present N , therefore

$$A \propto N \quad \text{and} \quad A_0 \propto N_0$$

Equation (3) can also be written as

$$A = A_0 e^{-\lambda t} \quad \dots\dots\dots (4)$$

where A is the activity of the sample of radioactive nuclei at time t and A_0 is the initial activity of the sample of radioactive nuclei at time $t = 0$.

5. Variation of Count Rate with time

Furthermore, since the count rate C of a sample of radioactive nuclei is a proportion of the activity A , we also have:

$$C \propto A \quad \text{and} \quad C_0 \propto A_0$$

Therefore, equation (4) can also be written as: $C = C_0 e^{-\lambda t}$

Half-life $t_{1/2}$

Another parameter that is useful in characterizing nuclear decay is the **half-life** $t_{1/2}$.

The half-life of a radioactive nuclide is the average time taken for half of the original number of nuclei in a sample of the radioactive nuclide to decay.

OR

The half-life of a radioactive nuclide is the average time taken for the activity of a sample of radioactive nuclide to halve.

The **above definition assumes a statistically large number of particles**, and the half-life represents an expectation over a long time.

- Now consider a radioisotope with a $t_{1/2} = 7$ years

Time, t	No of radioactive nuclei present, N	Simplifying
0	N_0	$\frac{N_0}{2^0}$
7 yr (1 $t_{1/2}$)	$\frac{N_0}{2}$	$\frac{N_0}{2^1}$
14 yr (2 $t_{1/2}$)	$\frac{N_0}{4}$	$\frac{N_0}{2^2}$
21 yr (3 $t_{1/2}$)	$\frac{N_0}{8}$	$\frac{N_0}{2^3}$



Nuclear half-life

Therefore, we see that after n half-lives, the time elapsed: $t = n \cdot t_{1/2}$,

the number of radioactive nuclei present in the sample: $N = \frac{N_0}{2^n}$

o Relationship between decay constant and half-life

From the definition of half-life, we can also see that:

$$\text{At } t = t_{1/2} \Rightarrow N = \frac{N_0}{2} \text{ and } N = N_0 e^{-\lambda t} \Rightarrow \frac{N_0}{2} = N_0 e^{-\lambda t_{1/2}} \Rightarrow t_{1/2} = \frac{\ln 2}{\lambda} \quad (\text{constants list})$$

Example 11: Radon Gas

Serway College Physics. Pg 948-Example 29.4

Radon $^{222}_{86}\text{Rn}$ is a radioactive gas that can be trapped in the basements of homes, and its presence in high concentrations is a known health hazard. Radon has a half-life of 3.83 days. A gas sample contains 4.00×10^8 radon atoms initially.

- How many atoms will remain after 14.0 days have passed if no more radon leaks in?
- What is the activity of the radon sample after 14.0 days?
- How long will it be before 99% of the sample has decayed?

a) no. of atoms remain, $N = N_0 e^{-\lambda t} = (4.00 \times 10^8) e^{-(\ln 2 / 3.83)(14.0)}$
 $= 3.17 \times 10^7$

b) Activity of radon sample after 14.0 days,
 $A = \lambda N$
 $= \left(\frac{\ln 2}{3.83 \times 24 \times 60 \times 60} \right) (3.17 \times 10^7) = 66.4 \text{ Bq}$

c) If 99% has decayed, 1% undecayed
 $\frac{N}{N_0} = e^{-\lambda t}$
 $0.01 = e^{-\lambda t}$

$$t = -\frac{1}{\lambda} \ln(0.01) = 25.4 \text{ days}$$



Lecture example solutions

18.6.3 Uses of Radioactive Nuclides

Radioactive nuclides are used in a myriad of applications from medicine to industry. Two interesting applications – one in medicine and the other in archaeology – are described below. Appendix C contains more applications of radioactive nuclides.

Medicine – Assessment of Thyroid Function

Iodine, which is absorbed into the blood stream, is collected by the thyroid gland for use in making thyroid hormones. Both over-active and underactive thyroids can lead to medical problems. To investigate the behaviour of the thyroid, tests are carried out to measure the accumulation of radioactive ^{131}I by the gland. The patient is given a tracer dose of dilute sodium iodide solution and the count rate from the thyroid is taken over a period of time and compared with that of a normal thyroid (Figure 21).

More recently, $^{99}\text{Tc}^m$ (an isotope of technetium) has been used instead of iodine. The patient is given sodium pertechnetate which is taken up in a similar way to iodine but is more easily released from the body. In addition, the half-life of $^{99}\text{Tc}^m$ is much shorter – ^{131}I is 8 days but $^{99}\text{Tc}^m$ is only 6 hours. The person thus remains radioactive for a much shorter period of time.

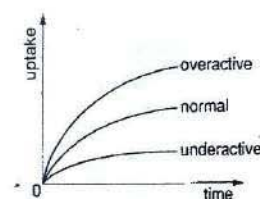


Figure 21 Assessment of Thyroid Function.

Archaeology – Carbon Dating

- Carbon-14 dating is a way of determining the age of certain archeological artifacts of a biological origin up to about 50,000 years old.

- Carbon isotopes that occur naturally are

$^{12}_6\text{C}$: stable, 99% abundance

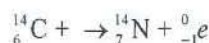
$^{13}_6\text{C}$: stable, 1% abundance

$^{14}_6\text{C}$: radioactive, $t_{1/2} = 5730$ year, trace amounts

- When cosmic rays from outer space collide with the upper atmosphere, one of the nuclear reactions that occur is the conversion of Nitrogen-14 to Carbon-14 by neutron capture :



The carbon-14 atoms then undergo beta-decay:



- Because $^{14}_6\text{C}$ is chemically identical to $^{12}_6\text{C}$, it can form carbon dioxide CO_2 . This radioactive CO_2 is absorbed by plants via photosynthesis. All living organisms that eat plant matter will also ingest this radioactive CO_2 .

- In the atmosphere, the rate of production of $^{14}_6\text{C}$ is approximately balanced by the rate of removal of $^{14}_6\text{C}$ and hence the ratio of $^{14}_6\text{C}$ to $^{12}_6\text{C}$ in the atmosphere is constant at approximately 1.3×10^{-12} .

The carbon atoms in all living organisms will correspondingly have this same ratio $^{14}_6\text{C} / ^{12}_6\text{C}$ as the organisms continuously exchange carbon dioxide with their surroundings.

- When an organism dies, the carbon-14 atoms decayed are not replenished, while the amount of carbon-12 remains constant in the organism. Thus the ratio $^{14}_6\text{C} / ^{12}_6\text{C}$ begins to decline.
- It is therefore possible to determine the time elapsed since death by measuring its $^{14}_6\text{C}$ activity.
- Using this technique, scientists have been able to identify samples of wood, charcoal, bone and shell as having lived from 1000 to 25000 years ago. This knowledge has helped us reconstruct the history of living organisms - including humans, during this time span.

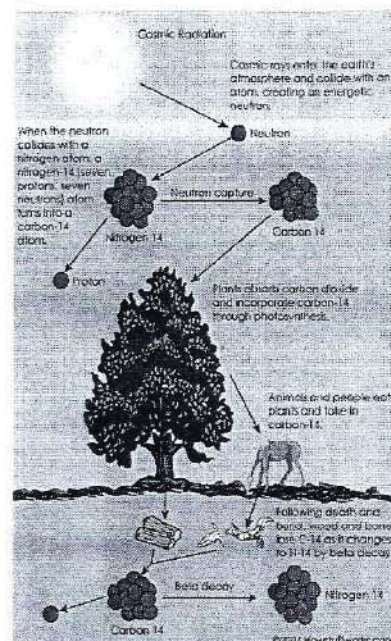


Figure 22. Carbon-14 Dating.



How carbon dating works?

Example 12: Carbon Dating

TYS N83/11/6 (part)

Living material contains a fairly constant small fraction of radioactive isotope $^{14}_6\text{C}$. After death, the activity from the $^{14}_6\text{C}$ present in the material decays with a half-life of 5.7×10^3 years. A certain quantity of carbon converted from recently dead organic matter to carbon dioxide and placed in a GM tube gave 12 counts per minute above background. If the same quantity of carbon were taken from a linen shroud and similarly treated, what count rate would you expect if the shroud were

- a relic dating from the early Christian era, about A.D. 33,
- a medieval forgery dating from about A.D. 1100?

a) For recently dead organic matter, i.e., $t=0$, $C_0 = (12 \text{ min}^{-1})$

Take current year as 1983

As shroud is from A.D. 33,

time that has passed, $t = 1983 - 33 = 1950 \text{ yrs}$

Using $C = C_0 e^{-\lambda t} = (12 \text{ min}^{-1}) e^{-\left(\frac{\ln 2}{5.7 \times 10^3}\right)(1950)} = 9.5 \text{ min}^{-1}$

b) As shroud is from A.D. 1100,

time that has passed,

$t = 1983 - 1100 = 883 \text{ yrs}$

Using $C = C_0 e^{-\lambda t}$

$= (12 \text{ min}^{-1}) e^{-\frac{\ln 2}{5.7 \times 10^3} \times 883}$

$= 10.8 \text{ min}^{-1}$



Lecture example solutions

18.6.4 Radiation Hazard

Radioactive emissions can cause damage to the cells in your body.

Most of this damage is due to **ionization** when the radiation passes, although if levels of radiation are high there can also be damage due to heating effects as your body absorbs the energy from the radiation.

○ Ionizing Radiation

Ionizing radiation encompasses energetic subatomic particles, ions or atoms moving at high speeds, and electromagnetic waves on the high-energy end of the electromagnetic spectrum. During an interaction with an atom, it removes electrons from the orbit of the atom, causing the atom to be electrically charged or ionized.

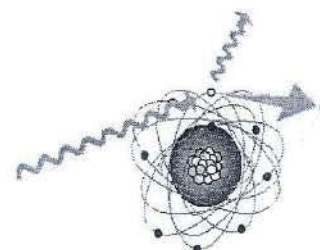


Figure 23. Incoming radiation interacting with a neutral atom, causing ionization.

○ Damage to Cells

The nucleus of a biological cell has been shown to be the part of the cell most vulnerable to radiation damage. This is because the nucleus contains the DNA molecules which carry the organism's genetic code.

○ Types of damage to cells

Damage may be **direct** or **indirect**.

1. Ionisation may cause **direct damage** to the DNA within the nucleus, when the whole DNA molecule may either be broken into fragments or have sections of it removed.
2. **Indirect damage** occurs when the radiation interacts with other molecules in the cell e.g. water, producing ions and radicals (H^+ , OH^- , $\text{H}\cdot$, $\text{OH}\cdot$) which can then attack other cell structures and DNA. They can also combine to form toxic substances like H_2O_2 which could initiate harmful reactions within the cells.

○ Effects of Damage to DNA

- **Cell death:** If the damage within the cell is great, then **cell death** occurs immediately or within a very short time.
- **Cell is unable to replicate:** The DNA in the cell nucleus may have been damaged so badly that the **cell is unable to replicate itself by cell division**. This may have serious long term consequences for the organism.
- **Cancer, genetic mutation and sterility:** At low doses, cellular reproduction may continue for many cell generations, giving rise to delayed effects such as **enhanced risk of developing cancer**. Rapidly multiplying cells are far more susceptible to damage and consequently, special attention is paid in the use of ionising radiation on unborn children and young persons. The reproductive system may also be affected, giving rise to an **increased risk of genetic mutations**. If the damage is severe enough, it can cause **sterility**.



How damaging is radiation?

Appendix A: Detecting Radiation

Nuclear radiation may come from the following:

1. Naturally occurring radioactive materials such as compounds of uranium, thorium and potassium.
2. From purified or man-made sources, which are stronger than the natural sources. The ones you find in the school laboratory are sealed and are called *closed* sources. They are stored in lead-lined boxes.
3. Space: High energy particles such as muons, pions, sigmas reach the earth's surface from space. All these decay into other particles very quickly. They can also be produced artificially in particle accelerators. There are various theories that these 'fundamental' particles (including protons, neutrons, electrons) are made from even more fundamental particles called quarks with a charge of $\pm \frac{1}{3}e$. It is unlikely that you will be able to detect these in the laboratory.

Wilson Cloud Chamber

A cloud chamber is a device that makes the paths of ionizing radiation entering it visible.

Water is introduced into chamber Y. If sufficient time is given, the air in the chamber becomes saturated with water vapour.

A series of adiabatic expansion is carried out to cool the air in chamber Y. As the air cools, the water condenses on the dust particles in air and will fall off as water droplets. After a series of controlled adiabatic expansions, the air could be made dust free and supersaturated i.e. the vapour pressure is greater than the saturated vapour pressure at the reduced temperature reached but no water vapour condenses. This supersaturated air is exposed to ionising radiation such as alpha, beta or gamma.

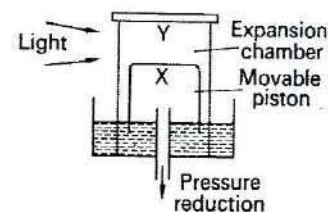


Figure 24.
Wilson Cloud Chamber

An ionising particle would collide with the air and water molecules to produce ions. This leaves a track of ions in its path which forms the nucleating centers for water to condense. Hence the expansions leave a track of water droplets showing the route the particle had taken. The light scattered from the tracks makes them visible.

Glaser's Bubble Chamber

In the same way as air can be super-saturated with water vapour, a liquid under pressure can be heated to a temperature higher than that at which boiling normally takes place and is then said to be superheated.

If the pressure is suddenly released, bubbles may not form in the liquid for a certain time interval. During this quiet period, if the liquid is exposed to ionising radiation, nucleating centers are obtained for bubble formation and the liquid quickly evaporates into bubble which grow rapidly, and the bubble track would reveal the path of the ionising radiation.

KIASU
ExamPaper



Cloud chamber

The Geiger-Müller (GM) Tube

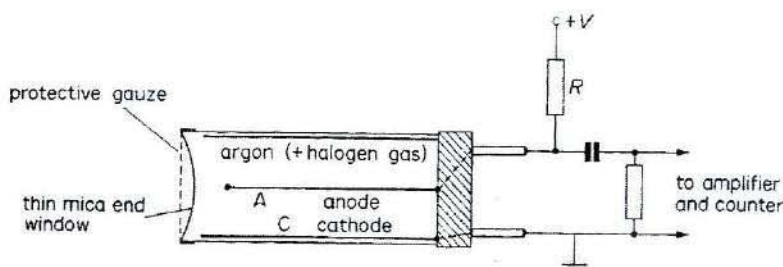


Figure 25. GM Tube

- A GM tube consists of a tube filled with a low-pressure inert gas such as helium, neon or argon, and an organic vapor or a halogen gas. The tube contains electrodes, between which there is a potential difference of several hundred volts, but no current flowing. The walls of the tube are either entirely metal or have their inside surface coated with a conductor to form the cathode while the anode is a wire passing up the center of the tube.
- When ionizing radiation passes through the tube, some of the gas molecules are ionized, creating positively charged ions, and electrons. The strong electric field created by the tube's electrodes accelerates the ions towards the cathode and the electrons towards the anode. The ion pairs gain sufficient energy to ionize further gas molecules through collisions on the way, creating an avalanche of charged particles. This results in a short, intense pulse of current which passes (or cascades) from the negative electrode to the positive electrode and is measured or counted.
- The usual form of GM tube is an *end-window tube*. This type is so-named because the tube has a window at one end through which ionizing radiation can easily penetrate. The other end normally has the electrical connectors. There are two types of end-window tubes: the glass-mantle type and the mica window type. The glass window type will not detect alpha radiation since it is unable to penetrate the glass, but is usually cheaper and will usually detect beta radiation and X-rays. The mica window type will detect alpha radiation but is more fragile.
- The ideal GM tube should produce a single current pulse on entry of a single ionising particle. It must not give any spurious pulses, and must recover quickly to the passive state. Unfortunately for these requirements, when positive argon ions reach the cathode and become neutral argon atoms again by obtaining electrons from it, the atoms can acquire their electrons in enhanced energy levels. These atoms then return to their ground state by emitting photons which can in turn produce further ionisation and hence cause spurious secondary pulse discharges. Some form of quenching of the ionisation is therefore essential. The disadvantage of quenching is that for a short time after a discharge pulse has occurred (the so-called *dead time*, which is typically a few microseconds), the tube is rendered insensitive and is thus temporarily unable to detect the arrival of any new ionising particle. This effectively causes a loss of counts at sufficiently-high count rates.
- External quenching uses control electronics to temporarily remove the high voltage between the electrodes. Self-quenching or internal-quenching tubes stop the discharge without external assistance, by means of the addition of a small amount of an organic vapor or a halogen such as bromine or chlorine. The choice of quencher molecules is important.
- If a poor gas quencher were introduced to the tube, the positive argon ions, during their motion toward the cathode, would have multiple collisions with the quencher gas molecules and transfer their charge and some energy to them. Neutral argon atoms would then be produced and the quencher gas ions would reach the cathode instead, gain electrons in excited states which would decay by photon emission, thereby producing spurious tube discharge as before. However, effective quencher molecules, when excited, do not lose their energy by photon emission but by dissociation into neutral quencher atoms. No spurious output pulses are then produced.

Counters – Scalers and Ratemeters

- Each ionising particle or radiation produces a pulse voltage in the external circuit of a GM detector. In order to measure the number of pulses from the detectors, some form of counter must be used.
- A scaler counts the actual number of pulses while a ratemeter gives directly the average number of pulses per second or **count rate**.

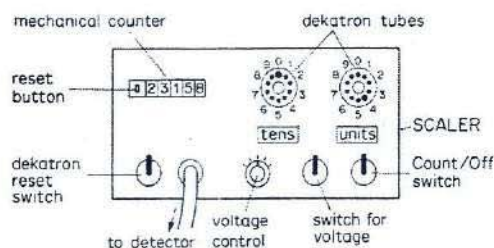


Figure 26. (a) A scaler

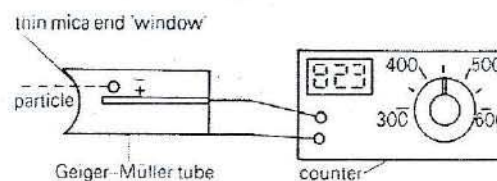


Figure 26. (b) Connection of a scaler to a G-M tube for detection of nuclear particles.

Scintillation Detector

Gas filled detectors like the GM tube are inefficient as gamma detectors as the probability of an interaction is low. A "solid" detector like the Scintillation detector addresses this shortfall. The incoming ionising radiation interacts with the scintillator to produce many electronic transitions. The decay of these transitions produces many photons of low energy i.e. the scintillator material fluoresces.

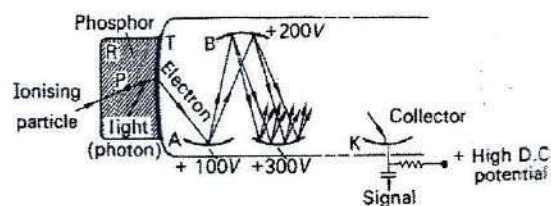


Figure 27.
Scintillation detector with photomultiplier

Each emitted photon travels through the scintillator material (transparent) and ejects an electron from a photosensitive surface. Each of these electrons goes on to emit more secondary electrons (the numbers of electrons are multiplied by a photomultiplier and turn into a voltage pulse so that pulse amplitude is a measure of the energy of the incident ionising radiation).

Appendix B: Background Radiation

Background radiation is the ionizing radiation that is constantly present in the environment. Sources of such radiation can be natural or man-made.

Sources of Radiation Exposure

From: NCRP Report No. 93

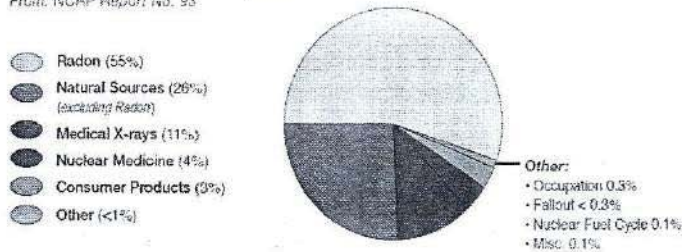


Figure 28. Sources of Radiation Exposure.

Natural Sources of Radiation

Naturally-occurring radiation accounts for approximately 80% of our exposure. Most of our exposure is from indoor radon, followed by radiation from outer space and from the Earth's crust.

o Radon

- Approximately 55% of our exposure to radiation comes from radon.
- Radon is a *colourless, odourless* and *relative inert* naturally occurring radioactive gas. Radon-222 is one of the elements in the long radioactive decay chain from uranium-238 (Figure 29), and the less common isotope radon-220 is part of the decay series from thorium-232 (Figure 30), both of which can be found in rocks and soil.

➤ Dangers of Radon Gas

Radon is the 2nd leading cause of lung cancer in the United States, with only smoking causing more lung cancer deaths.

Key dangers associated with radon gas is not so much of radon itself but its radioactive daughters which are solid particles that easily attach themselves to dust particles which are inhaled.

When inhaled these particles stick to the surfaces of the bronchial tubes and the α -particles emitted in the decay of the radon daughters damage the cells they are attached to, causing cancer.

➤ How does Radon get into the homes?

Radon usually migrates into homes and other buildings from the ground through openings in the floor or wall. Buildings trap radon inside, where it accumulates and may become a health hazard. Any home or building may have a radon problem, including new and old homes, well-sealed and drafty homes, and homes with or without basements.

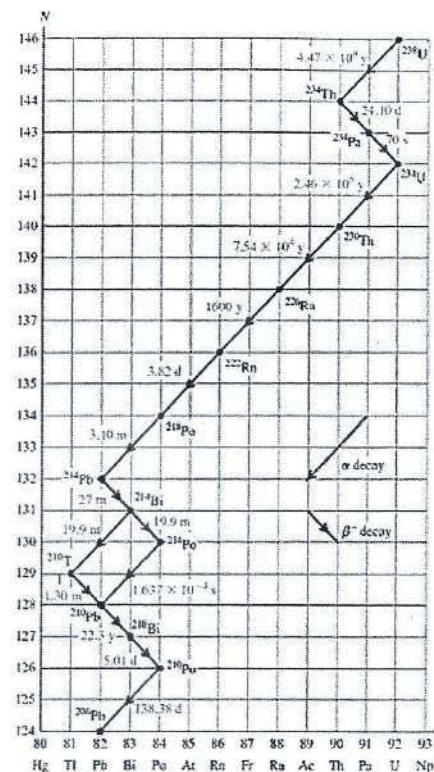


Figure 29. The ^{238}U decay series, terminating with stable nuclide ^{206}Pb .

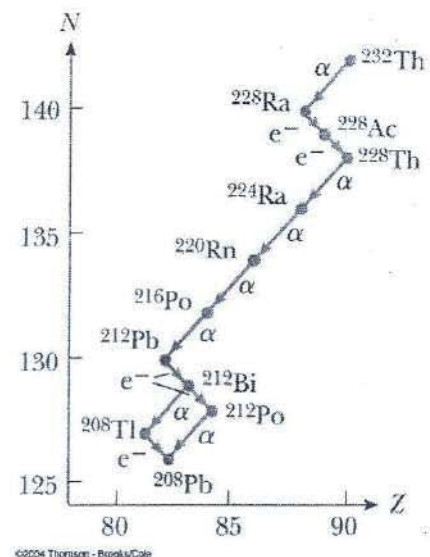


Figure 30. The ^{232}Th Series

○ Soil and Buildings

- Natural radioactive material in rocks and soil account for about another 8% of the radiation dose a person typically receives in a year.
- The earth's crust contains small amounts of uranium, thorium, and radium as well as radioactive isotopes of several elements including potassium.
- The radiation dose also comes from some building materials (such as bricks and concrete).

○ Cosmic Radiation

Another 8% of our exposure comes from cosmic radiation

- Cosmic radiation mainly originates from the sun and outer space. Cosmic radioactive nuclides (mainly carbon-14), are produced through interactions of the cosmic rays with atoms in the atmosphere.
- Cosmic rays are extremely energetic particles, primarily protons, which originate in the sun, other stars, and from violent cataclysms in the far reaches of space. Cosmic ray particles interact with the upper atmosphere of the earth and produce showers of lower energy particles. Many of these lower energy particles are absorbed by the earth's atmosphere.
- The cosmic radiation exposure varies with altitude. The exposure of an individual to cosmic rays is greater at higher elevations than at sea level. For example, a passenger in a jetliner traveling at 37,000 feet would receive about 60 times as much cosmic radiation as would a person standing at sea level for the same length of time.

○ Natural Radioactivity in the Body

- Small traces of many naturally occurring radioactive materials are present in the human body. These come mainly from naturally occurring radioactive nuclides present in the food we eat and in the air we breathe. These isotopes include tritium (^3H), carbon-14, and potassium-40.

Man-Made Sources

Since the discovery of radiation, people have benefited from the use of radiation in medicine and industry. Man-made sources of radiation account for about 20% of our total exposure to radiation.

○ Medicine

- The radiation used in medicine is the largest source of man-made radiation to which people in most countries are exposed.
- Most of the exposure is from diagnostic X-rays. Physicians use X-rays in more than half of all medical diagnoses to determine the extent of disease or physical injury.
- Radiation is also used in cancer treatments, where precisely targeted radiation destroys diseased cells without killing nearby healthy cells.
- Radiopharmaceuticals, another medical treatment, are used to locate tumors in a patient's body and to treat cancer. One-third of all successful cancer treatments involve radiation.

○ Consumer Products

- Small amounts of man-made background radiation is attributable to consumer products like colour televisions, smoke detectors, gas lantern mantles, natural gas heating and cooking fuel, and mining and agriculture products, such as coal, granite and potassium salt.

- **Nuclear Power Plants**

- Nuclear power reactors, which use uranium - a radioactive isotope, reduce reliance on fossil fuels. Nuclear power plant operations account for less than a hundredth of a percent of the average American's total radiation exposure.

- **Nuclear Fallouts**

- Nuclear weapons derive their explosive power from the uncontrolled radioactive break-up of plutonium and uranium. This yields a large number of radioactive daughter products that are blown high into the atmosphere and are carried around the earth. These radioactive elements gradually fall back (fall-out) to earth over a period of many years.
- During the 1950s and early 1960s, many test explosions were carried out in the atmosphere. In 1963 an Atmospheric Test Ban Treaty was signed and most subsequent tests have been conducted underground.
- The peak year for radioactive fallout was 1963 and levels have been declining since then. Levels of fallout are now less than ten per cent of what they were at their peak.



Who could survive
nuclear bomb?

- **Tobacco**

- The tobacco in cigarettes contains Pb-210.
- Pb-210 is a naturally occurring radionuclide that precipitates out of the atmosphere and deposits on the leaves of tobacco.
- When the tobacco is inhaled, the smoker receives a dose from the inhaled Pb-210 as well as Po-210, the decay product of Pb-210. Pb-210 is deposited on the surfaces of bones and Po-210 is deposited in the liver, kidney and spleen.

Appendix C: Uses of Radioactive Nuclides

Radioactive nuclides are used in some industries and in nuclear medicine for their chemical properties and as sources of radiation. Some examples are listed below.

o In Medicine:

- As a tracer. Any radioactive nuclide which is to be used as a tracer for diagnostic purposes must be incorporated into a pharmaceutical product which can be administered to the patient.

The pharmaceutical produce and the radioactive nuclide which it incorporates should have the following properties:

- It should concentrate in the organ or system which is to be imaged or assessed.
- It should not change the functioning of the organ, i.e. the pharmaceutical behaves in the same way as the substance it is replacing.
- The half-life of the radioactive nuclide should be as short as possible to reduce radiation damage, but sufficiently long to enable measurements to be made.

Some examples of the use of radioactive tracers in Medicine is listed below:

- o **Thrombosis (Blood clots).** A radioactive tracer, iodinated fibrinogen is introduced into the blood. If there is a blood clot forming in some part of the body, a concentration of the radioactive fibrinogen will occur at that point. The concentration can be located using a radiation detector.
- o **Haemorrhage Inside the body.** Similar to thrombosis, the radioactive tracers are used to locate the precise site of haemorrhage in an emergency situation. Radioactive chromium is injected into the body and is taken up by the red blood cells and carried uniformly throughout the body. However, there will be more blood accumulating at the haemorrhage site and hence a greater activity which can be located using a radiation detector.
- Cancer treatment. Immature cells and cells that are growing or dividing most rapidly are most sensitive to radiation. This is made use of in radiation treatment of cancer.

Often, cancer cells are growing rapidly and are therefore much more likely to be killed by a high dose of gamma radiation from cobalt-60 source than are normal cells that divide less frequently.



What is nuclear medicine?

o In Industry:

- As a tracer. Tracers are used for detecting leaks in underground pipes by doping the water in the pipes and using a detector to check where the water flows. Tracers are also used to check the wear of a machine part. For example, we can check the wear of the piston if it has been doped with a radioactive material during manufacture by checking the activity in the oil that is used with the piston.
- To check for cracks and imperfections
 - o Gamma rays can be used in place of X-rays to photograph solid objects. Gamma rays are directed onto the object and its image is captured on a photographic film placed on the other side. The film can reveal any imperfections e.g. a crack. The technique has an advantage over X-rays as no high voltage electrical sources are needed.
 - o The same technique can also be used to check and control the thickness of paper or plastic during the manufacturing process. The thickness of rolled sheets of plastic can be controlled by placing a beta source on top of the sheet and a detector beneath it. If the sheet is too thin, the count rate will rise. This rise in count rate can then be directly coupled to the rollers so that they separate slightly.

o **In Agriculture:**

- Suppose a plant research centre wishes to investigate the ability of a particular plant to take up magnesium through its roots, it is possible to water the plant with a fertilizer containing a known, extremely low concentration of radioactive magnesium-28.

If a detector is placed in contact with the leaves of the plant, an increase in count rate soon after watering in the fertilizer indicates that the magnesium has travelled through the plant to the leaves.

This can happen within a time as short as 20 min.

o **As a sterilization tool:**

- Radiation can also be used as a means of preserving food because exposure to high levels of radiation can destroy or incapacitate bacteria and mould spores.

Techniques include exposing foods to gamma rays, high-energy electron beams, or x-rays.

Food preserved this way can be placed in a sealed container (to keep out new spoiling agents) and stored for long periods of time.

Appendix D: Calculating Radiation Doses

- **Radiation dosimetry** is the quantitative description of the effect of radiation on living tissue.
- **Absorbed Dose:**
 - Absorbed dose of radiation is defined as the energy delivered to the tissue per unit mass.
 - The SI unit of absorbed dose is *gray* (Gy):
 - $1 \text{ Gy} = 1 \text{ J/kg}$
 - Another commonly used unit is the *rad*.
 - $1 \text{ rad} = 0.01 \text{ J/kg} = 0.01 \text{ Gy}$
- **Relative Biological Effectiveness (RBE):**
 - The absorbed dose by itself is not an adequate measure of biological effect because equal energies of different radiation causes different extent of biological effect.
 - This variation is described by a numerical factor called the **relative biological effectiveness (RBE)** or the **quality factor (QF)** of each specific radiation.
 - The table below shows some approximate values of RBE for several types of radiation. These values are obtained from experimental observation. All these values depend somewhat on the kind of tissue in which the radiation is absorbed and on the energy of the radiation.

RBE for Different Types of Radiation

Radiation	RBE / (Sv/Gy) or (rem/rad)
X-rays and γ -rays	1
Electrons, β -particles	1.0 -2.0
Slow neutrons	3 - 5
Fast neutron, protons, singly charged particles	10
α -particles, multiply charged particles	20

- **Biologically Equivalent Dose :**
 - The biological effect is described by the quantity **biological equivalent dose**, or simply equivalent dose.
 - The **equivalent dose** is defined as the product of the absorbed dose and RBE of the radiation received:

$$\text{Equivalent Dose} = \text{RBE} \times \text{Absorbed dose}$$

- The SI unit of equivalent dose for humans is the **Sievert (Sv)**:

$$1 \text{ Sv} = \text{RBE} \times 1 \text{ Gy}$$
- Another common unit, used in some countries, is the **rem (röntgen equivalent for man)**:

$$1 \text{ rem} = \text{RBE} \times 1 \text{ rad}$$